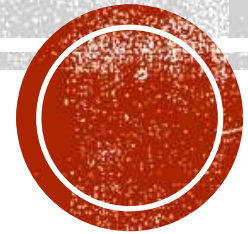
SCRIPPS INSTITUTION OF OCEANOGRAPHY

CAUSAL MEDIATION ANALYSIS AND APPLICATIONS IN ENVIRONMENTAL EPIDEMIOLOGY



18 September 2022

Chen Chen & Tarik Benmarhnia (UC San Diego)

ISEE, 34th Annual Conference

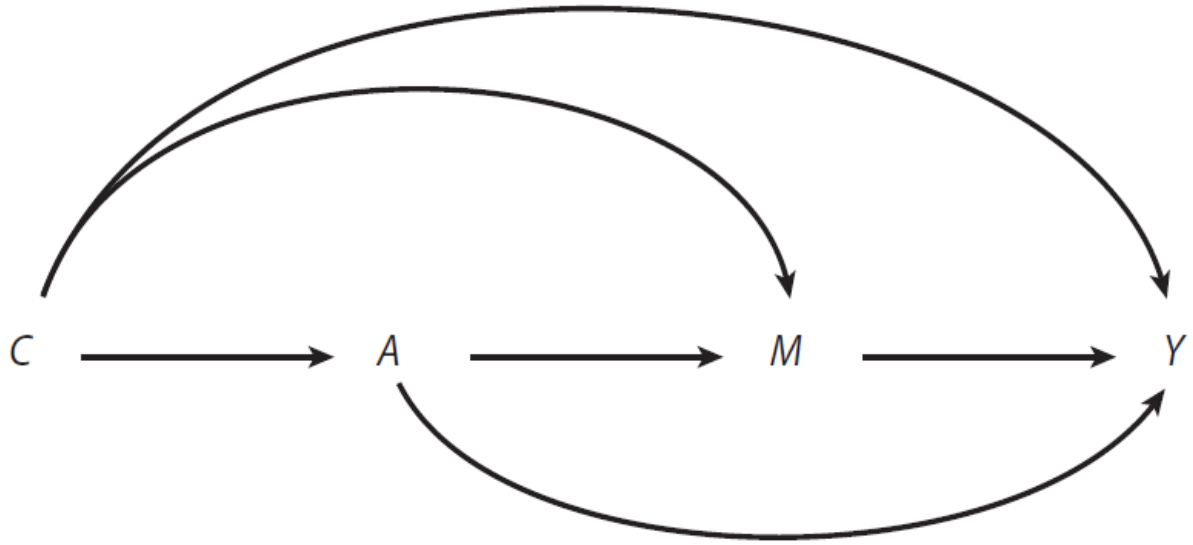
Athens, Greece

ISEE 2022 ATHENS-GREECE

OUTLINE

- **Part I**: Mediation analysis in epidemiology
  - Motivation, notation and estimands
  - 2-way decomposition with one mediator
  - Multiple Mediators
  - Unifying Mediation and Interaction: 4-way decomposition
  - Mediation analysis for environmental justice studies
- **Part II** – Mediation Analysis in Practice
- **Part III** - Summary and advanced topics

INTUITION



MULTIPLE MEDIATION ANALYSES APPLICATIONS

ETIOLOGIC RESEARCH:
UNDERSTANDING THE MECHANISMS

The role of cardiovascular disease in the relationship between air pollution and incident dementia: a population-based cohort study

Sindana D Ilango,[1,2*] Hong Chen,[3,4,5,6] Perry Hystad,[7] Aaron van Donkelaar,[8] Jeffrey C Kwong,[4,5,6,9,10] Karen Tu,[9,10] Randall V Martin[8,11] and Tarik Benmarhnia[2,12]

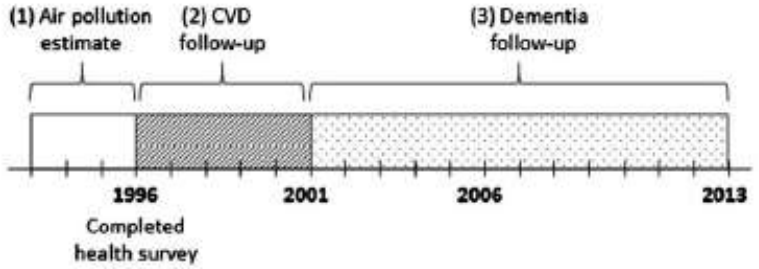


Table 3. Total, controlled direct and natural indirect effects of ambient air pollutant through cardiovascular disease (Cox proportional hazards model)

Pollutant	Total effect HR[a] (95% CI)	Natural direct effect HR[b] (95% CI)	Natural indirect effect HR[b] (95% CI)
NO_2[c]	1.10 (0.99–1.19)	1.09 (1.00–1.18)	1.01 (0.98–1.03)
$PM_{2.5}$[c]	1.29 (0.99–1.64)	1.22 (0.95–1.56)	1.06 (0.99–1.12)

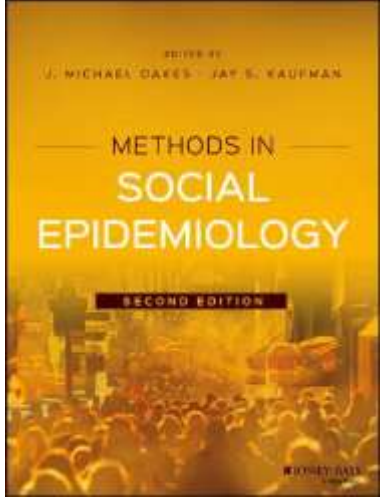
Table 4. Total, controlled direct and natural indirect effects of ambient air pollutant through cardiovascular disease (Aalen additive hazards model)

Pollutant	Total effect Estimate[a] (95% CI)	Natural direct effect Estimate[b] (95% CI)	Natural indirect effect Estimate[b] (95% CI)
NO_2[c]	100×10^{-5} (1.20×10^{-5}–100×10^{-5})	100×10^{-5} ($<100 \times 10^{-5}$–100×10^{-5})	1.45×10^{-5} (1.00×10^{-5}–2.56×10^{-5})
$PM_{2.5}$[c]	100×10^{-5} (3.60×10^{-5}–300×10^{-5})	100×10^{-5} ($<100 \times 10^{-5}$, –300×10^{-5})	4.20×10^{-5} (2.85×10^{-5}–6.91×10^{-5})

[a]Total effect obtained from product method.
[b]Adjusted for age, sex, education, marital status, income quintile, smoking status, body mass index, physical activity, rural residence and northern region; area level: recent immigrants, unemployment and education.
[c]NO_2 per 5 ppb, $PM_{2.5}$ per 10 $\mu g/m^3$.

DISPARITY RESEARCH: UNDERSTANDING THE DRIVERS OF HEALTH INEQUALITIES

- Mediation of the association between Racial ethnical status (A) and infant mortality (Y) by breastfeeding prior to hospital discharge (M).
  - Naimi et al. (2016). Mediation Analysis for Health Disparities Research. *American Journal of Epidemiology*.
- Benmarhnia, T., Hajat, A., & Kaufman, J. S. (2021). Inferential challenges when assessing racial/ethnic health disparities in environmental research. Environmental Health, 20(1), 1-10.
- Mediation of the association between racial /ethnical status (A) and intima-media thickness (Y) by Air pollution Exposure (M).
  - Jones, M. R (2015). Ambient air pollution and racial/ethnic differences in carotid intima-media thickness in the Multi-Ethnic Study of Atherosclerosis (MESA). *Journal of epidemiology and community health*, jech-2015.



See Chap. 16

Oakes, J, and Kaufman J, eds. Methods in social epidemiology. John Wiley & Sons, 2017.

INTERVENTION RESEARCH

- Keele, L., Tingley, D., & Yamamoto, T. (2015). Identifying mechanisms behind policy interventions via causal mediation analysis. *Journal of Policy Analysis and Management*, *34*(4), 937-963.

- Mediation of the association between Air Quality Regulation (A) and Hospitalizations (Y) by local emissions and economic activity (M).
  - Zigler et al. (2016). Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. HEI research Report: https://www.healtheffects.org/publication/causal-inference-methods-estimating-long-term-health-effects-air-quality-regulations

AN INCREASING USE IN ENVIRONMENTAL EPIDEMIOLOGY

The mediation effect of placental weight change in the association between prenatal exposure to selenium and birth weight

Evidence from a prospective birth cohort study in China

Jiaqi Wang[a,b], Rui Qian[c], Yiding Wang[d], Moran Dong[a,e], Xin Liu[a], He Zhou[a,b], Yufeng Ye[f], Guimin Chen[a,g], Dengzhou Chen[a,b], Lixia Yuan[a,g], Jianpeng Xiao[a], Guanhao He[a], Jianxiong Hu[a], Weilin Zeng[a], Zuhua Rong[a], Qianqian Zhang[d], Mengya Zhou[d], Juan Jin[b], Jingjie Fan[h], Jiufeng Sun[b], Wenjun Ma[a,i], Bo Zhang[d,e], and Tao Liu[a,j]

Mediation of the Relationship between Maternal Phthalate Exposure and Preterm Birth by Oxidative Stress with Repeated Measurements across Pregnancy

Kelly K. Ferguson,[1] Yin-Hsiu Chen,[2] Tyler J. VanderWeele,[3] Thomas F. McElrath,[4] John D. Meeker,[1] and Bhramar Mukherjee[2]

Air Pollution and Adverse Pregnancy and Birth Outcomes: Mediation Analysis Using Metabolomic Profiles

Kosuke Inoue[1] · Qi Yan[1] · Onyebuchi A. Arah[1,2,3] · Kimberly Paul[1] · Douglas I. Walker[4] · Dean P. Jones[5,6] · Beate Ritz[1,7,8]

The mediating role of lung function on air pollution-induced cardiopulmonary mortality in elderly women: The SALIA cohort study with 22-year mortality follow-up

Andrea Dalecká[b,c], Claudia Wigmann[a], Sara Kress[a], Hicran Altug[a], Vítězslav Jiřík[b,c], Joachim Heinrich[d], Michael J. Abramson[e], Tamara Schikowski[a,*]

Unconventional natural gas development and adverse birth outcomes in Pennsylvania: The potential mediating role of antenatal anxiety and depression

Joan A. Casey[a,b,*], Dana E. Goin[c], Kara E. Rudolph[d], Brian S. Schwartz[e,f,g], Dione Mercer[f], Holly Elser[c], Ellen A. Eisen[a], Rachel Morello-Frosch[h,i]

DNA methylation as a mediator of associations between the environment and chronic diseases: A scoping review on application of mediation analysis

Ryosuke Fujii[a,†], Shuntaro Sato[b,†], Yoshiki Tsuboi[a], Andres Cardenas[c], and Koji Suzuki[a]

DIFFERENT APPROACHES

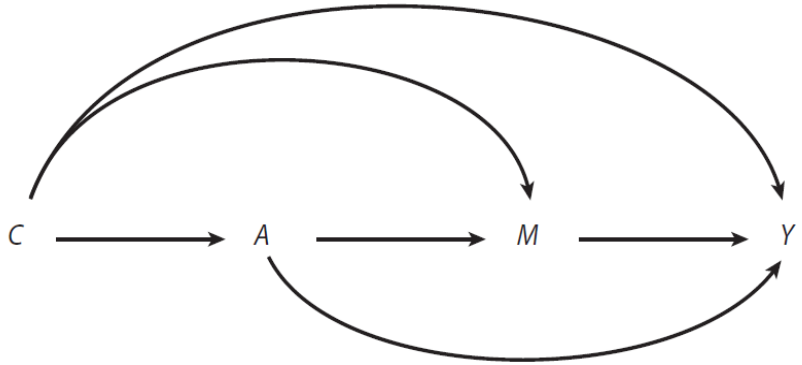
- The Difference method
- Structural Equation Modelling
  - VanderWeele, Tyler J. "Invited commentary: structural equation models and epidemiologic analysis." American journal of epidemiology 176.7 (2012): 608-612.

Not recommended

- **The Product Method**
  - Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol.1986;51(6):1173–1182.
  - VanderWeele, Tyler J. "Mediation analysis: a practitioner's guide." Annual review of public health 37 (2016): 17-32.
- **Weighting approaches**
  - VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology (Cambridge, Mass.), 25(2), 300.

THE PRODUCT METHOD

- Approach proposed by Baron & Kenny, 1986 to decompose the total effect into direct and indirect effects
- Augmented with modern approaches based on the counterfactual framework and different assumptions
  - Robins and Greenland, 1992; Pearl 2001; Kaufman et al. 2004; VanderWeele 2015



VanderWeele, 2016

COUNTERFACTUAL APPROACH TO MEDIATION ANALYSIS

- The counterfactual approach clarifies the assumptions needed to identify direct and indirect effects
- Framework and Notation
  - Let A be a treatment, M be a mediator, Y be an outcome and Cn potential confounders
  - Let Y (a,m) be the potential outcome Y when intervening to set A to a and M to m
  - Let M(a) be the potential outcome M when intervening to set A to a

ESTIMANDS AND DEFINITIONS

- **Total effect**
- **Natural Direct Effect (NDE)**
  - Equivalent to the Control Direct Effect (CDE) in the absence of exposure-mediator interaction
  - Corresponds to the effect of A on Y, after intervening to fix the level of M to a certain value *m*
- **Natural indirect effect (NIE)**
  - Corresponds to the change in Y when the level of A is held constant (A=a), and M changes to what it would have taken for one unit increase in A (A=a*).

ANALYTICAL APPROACH
TWO REGRESSIONS, ONE PRODUCT

- First linear regression:

$$E[Y \mid a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 c$$

- Second linear regression:

$$E[M \mid a,c] = \beta_0 + \beta_1 a + \beta_2 c$$

- Then:
  - Natural Direct effect: θ_1
  - Natural Indirect effect will be: $\beta_1 \theta_2$
  - The total effect will be: $\theta_1 + (\beta_1 \theta_2)$

FOR BINARY OUTCOMES

CONTINUOUS MEDIATOR, BINARY EXPOSURE, NO INTERACTION RARE DISEASE ASSUMPTION

- First logistic regression:

$$\text{Logit}\{P(Y = 1 \mid a,m,c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 c$$

- Second linear regression:

$$E[M \mid a,c] = \beta_0 + \beta_1 a + \beta_2 c$$

- Then:
  - $OR^{CDE} = \exp\{\theta_1 (a-a^*)\}$
  - $OR^{NIE} = \exp\{(\theta_2 \beta_1 (a-a^*)\}$
  - Proportion Mediated Measure = $OR^{CDE}(OR^{NIE}-1) / (OR^{CDE}OR^{NIE} - 1)$

FOR COUNT OUTCOMES

CONTINUOUS MEDIATOR, BINARY EXPOSURE, NO INTERACTION

- First Poisson regression:

$$E[Y \mid a,m,c] = \exp(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 c)$$

- Second linear regression:

$$E[M \mid a,c] = \beta_0 + \beta_1 a + \beta_2 c$$

- Then:
  - $RR^{CDE} = \exp\{\theta_1 (a-a^*)\}$
  - $RR^{NIE} = \exp\{(\theta_2 \beta_1 (a-a^*)\}$
  - Proportion Mediated Measure = $RR^{CDE}(RR^{NIE}-1) / (RR^{NDE}RR^{NIE} - 1)$

ASSUMPTIONS IN MEDIATION ANALYSES

- We will consider 4 different assumptions that are compulsory to estimate causal effects.
  - From VanderWeele 2015; *Explanation in Causal Inference: Methods for Mediation and Interaction*

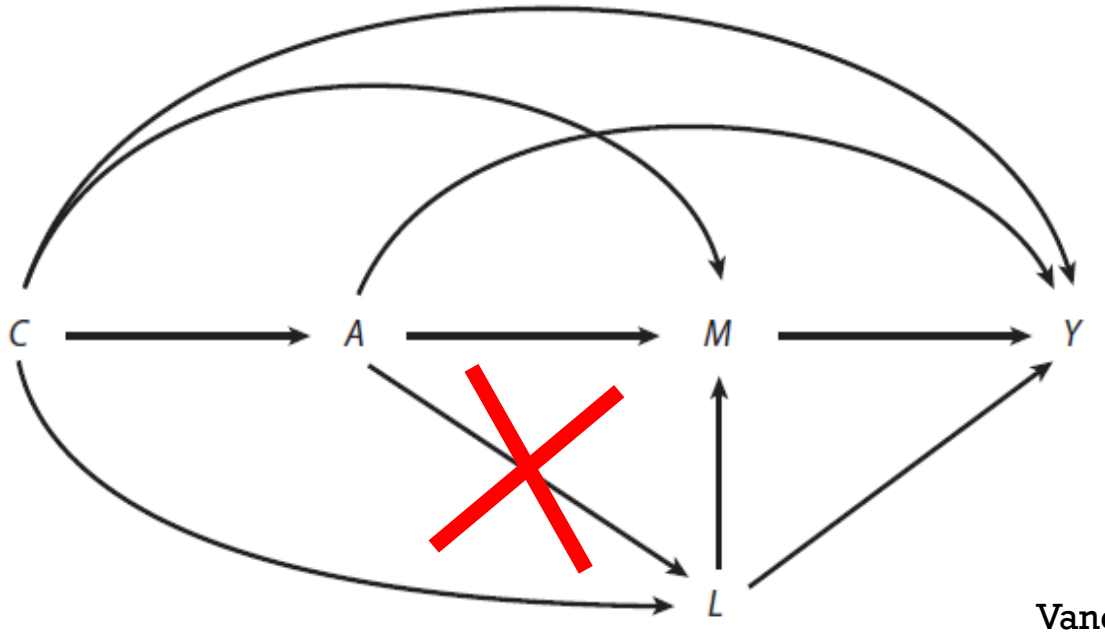
- A1: Exposure – Outcome confounding
- A2: Mediator – Outcome confounding
- A3: Exposure – Mediator confounding

Must be controlled and we assume NO unmeasured confounders

Sensitivity analyses can be useful

- A4: NO mediator – Outcome confounder itself affected by the exposure

ASSUMPTIONS IN MEDIATION ANALYSES



VanderWeele, 2016

CASE STUDY #1

THE ROLE OF OZONE AS A MEDIATOR IN THE RELATION BETWEEN HEAT WAVES AND MORTALITY

Alari A, Schwarz L, Chen C, Hansen K, Chaix B, Benmarhnia T. *The role of ozone as a mediator in the relation between heat waves and mortality in 15 French urban agglomerations*. American Journal of Epidemiology [In Press]

THE LINKS BETWEEN EXTREME HEAT AND OZONE IN RELATION TO POPULATION HEALTH

- Extreme heat and ozone have been shown to impact respiratory and CVD outcomes
- Ozone is a mediator in the relationship between heat and health
  - Ozone is generated by some precursors (NOx, NO2..), sunlight and <u>when temperatures are high</u>
  - Multiple papers have proposed such mechanism, but little empirical evidence to date

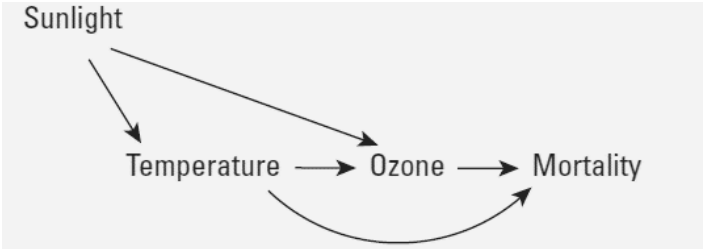
Does Air Pollution Confound Studies of Temperature?

Jessie P. Buckley,[a] Jonathan M. Samet,[b] and David B. Richardson[a]

The Role of Ambient Ozone in Epidemiologic Studies of Heat-Related Mortality

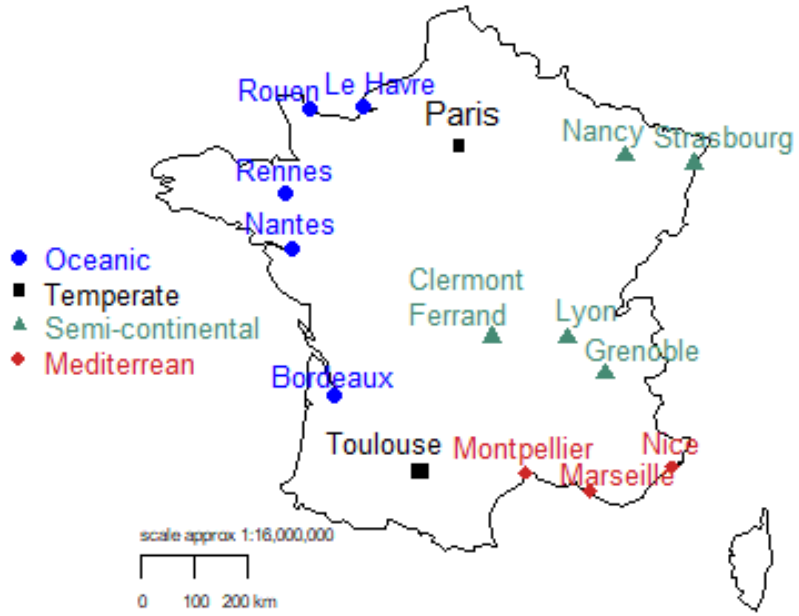
Colleen E. Reid,[1] Jonathan M. Snowden,[2,3] Caitlin Kontgis,[4] and Ira B. Tager[2]

[1]Department of Environmental Health Sciences, and [2]Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, USA; [3]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, Oregon, USA; [4]Department of Geography, Center for Sustainability and the Global Environment, University of Wisconsin–Madison, Madison, Wisconsin, USA

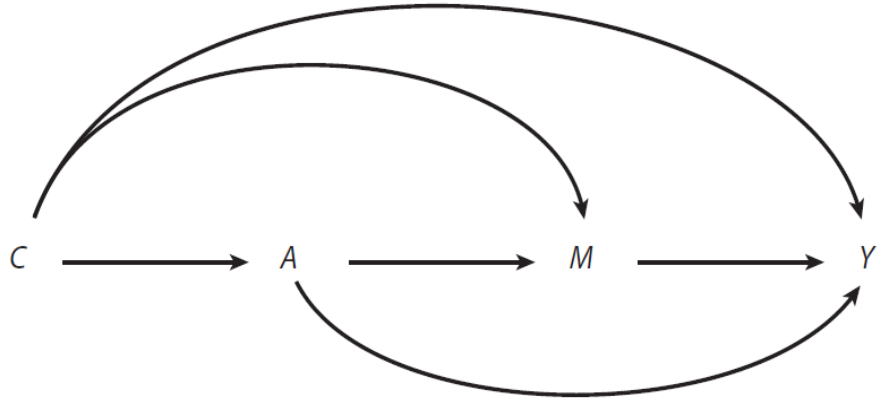


THE ROLE OF OZONE AS A MEDIATOR IN THE RELATION BETWEEN HEAT WAVES AND MORTALITY IN 15 FRENCH URBAN AGGLOMERATIONS

- Aims of this paper:
  - To decompose the total effect of heat wave on mortality into natural direct and indirect effects via increasing ozone levels
  - Compare the proportion mediation across French agglomerations
- Data Sources
  - 15 major French urban agglomerations, summer period (years 2000 to 2015)
  - We used the Official Météo-France Heat Wave definition for each city
    - We also analyzed several alternative HW definitions
  - Ozone and No2 obtained from air pollution monitors in each city
  - Mortality: daily counts for non-accidental deaths
    - We also considered respiratory and CVD deaths



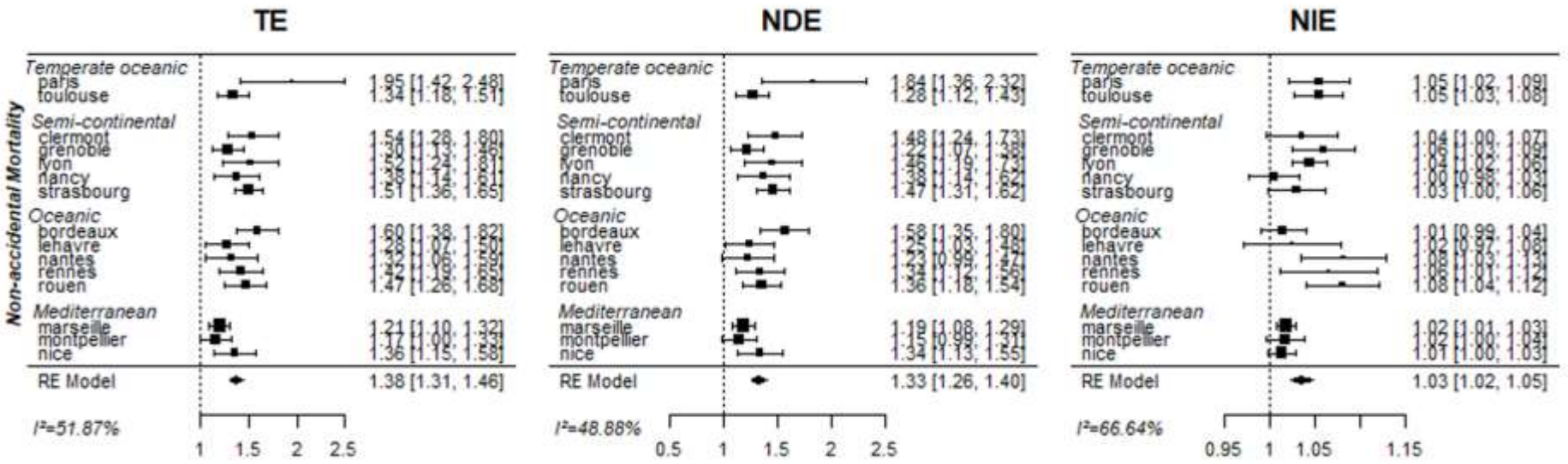
ANALYTICAL APPROACH



- We conducted a causal mediation analysis through a regression-based approach coupled with DLM
  - We estimated total effects, natural direct and indirect effects
  - With 4 identification assumptions
  - We did not identify any E-M interaction, so NDE and CDE were similar
- Two sequential regressions
  - $E[Y|hw, o3, c] = exp(\theta_0 + \theta_1 hw + \theta_2 o3 + \sum \theta_n c_n)$
  - $E[o3|hw, c] = \beta_0 + \beta_1 hw + \sum \beta_n c$
- *Then*
  - $NDE = \exp(\theta_1)$
  - $NIE = \exp(\theta_2 \beta_1)$
  - $PM = RR_{NDE}(RR_{NIE} - 1)/(RR_{NDE} * RR_{NIE} - 1)$

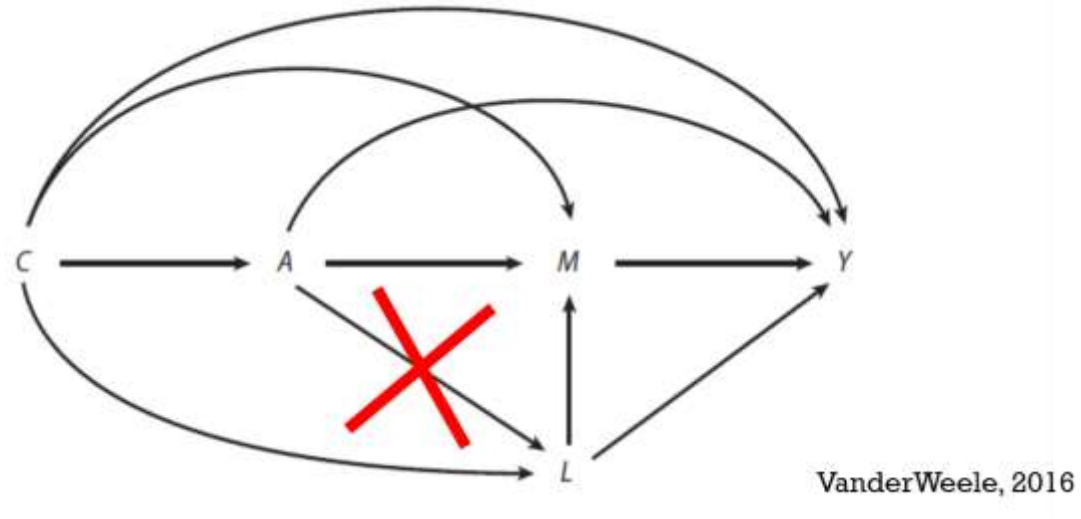
https://github.com/aalari/Heat-wave_O3_Mediation

MAIN RESULTS FOR NON-ACCIDENTAL MORTALITY



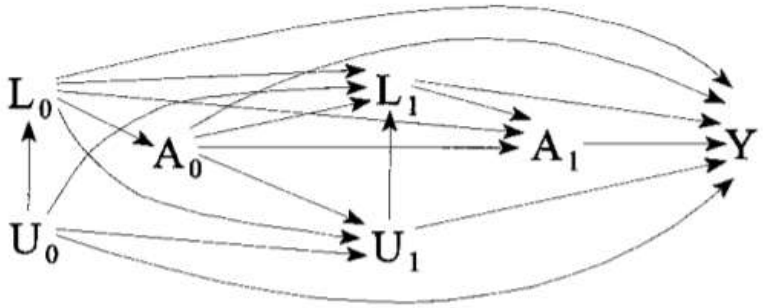
Urban Agglomeration	
Temperate-Oceanic climate	**PM (%)**
Paris	11%
Toulouse	20%
Semi-continental climate	
Clermont-Ferrand	10%
Grenoble	25%
Lyon	13%
Nancy	2%
Strasbourg	8%
Oceanic climate	
Bordeaux	4%
Le-Havre	11%
Nantes	31%
Rennes	20%
Rouen	23%
Mediterranean climate	
Marseille	10%
Montpellier	11%
Nice	5%

ASSUMPTION 4 VIOLATION



VanderWeele, 2016

- The analogy with another common issue in epidemiology: time-varying exposures and confounding
- Potentially a problem with multiple mediators too:
  - This is because *M* (1) may affect both *M* (2) and *Y* and thus be a mediator-outcome confounder.



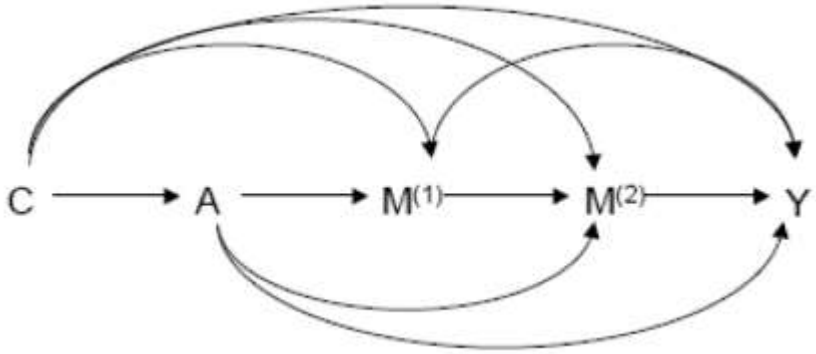
A SIDE NOTE ON TREATMENT CONFOUNDER FEEDBACK, G-METHODS AND INVERSE PROBABILITY OF TREATMENT WEIGHTING (IPTW)

[SEE ADDITIONAL SLIDES]

MULTIPLE MEDIATORS

MULTIPLE MEDIATORS

- When we are interested by decomposing the Total Effect into more than 1 indirect pathway.
- Vector of mediators **M** = (M (1), …, M (K))
- **One important question is: are the mediators connected?**



A REGRESSION-BASED APPROACH FOR MULTIPLE MEDIATORS (NOT AFFECTING EACH OTHER)

- We need to fit the following models:

$$\begin{aligned} E[Y|a,\mathbf{m},c] &= \theta_0+\theta_1 a+\theta_2^{(1)} m(1)+\theta_2^{(2)} m^{(2)}+\ldots+\theta_2^{(K)} m^{(K)}+\theta_4' c \\ E\left[M^{(i)}|a,c\right] &= \beta_0^{(i)}+\beta_1^{(i)} a+\beta_2^{(i)'} \quad for \quad i=1,\ldots,K. \end{aligned}$$

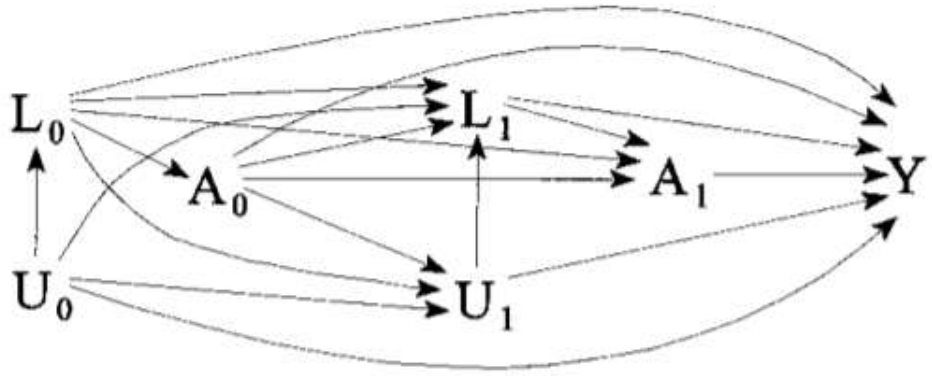
- CDE and NDE and NIE are then given by

$$\begin{aligned} E[Y_{a\mathbf{m}}-Y_{a^*\mathbf{m}}|c] &= \theta_1(a-a^*) \\ E\left[Y_{a\mathbf{M}_{a^*}}-Y_{a^*\mathbf{M}_{a^*}}|c\right] &= \theta_1(a-a^*) \\ E\left[Y_{a\mathbf{M}_a}-Y_{a\mathbf{M}_{a^*}}|c\right] &= \left[\beta_1^{(1)}\theta_2^{(1)}+\ldots+\beta_1^{(K)}\theta_2^{(K)}\right](a-a^*) \end{aligned}$$

- CDE are simply the coefficient for the exposure θ1 that contains all of the mediators
- The NIE is simply $\sum_{i=1}^{k}\theta_1\times\beta^k$

POTENTIAL ASSUMPTION VIOLATION

- When the mediators affect one another, the approach of handling one mediator at a time suffers **from another difficulty**
- Indeed, for the second (and potentially each subsequent) mediator, assumption A.4 will not hold if the mediators are considered one at a time.
- This is because *M* (1) may affect both *M* (2) and *Y* and thus be a mediator-outcome confounder.
- Similar to a time varying exposure situation



A REGRESSION-BASED APPROACH FOR MULTIPLE MEDIATORS

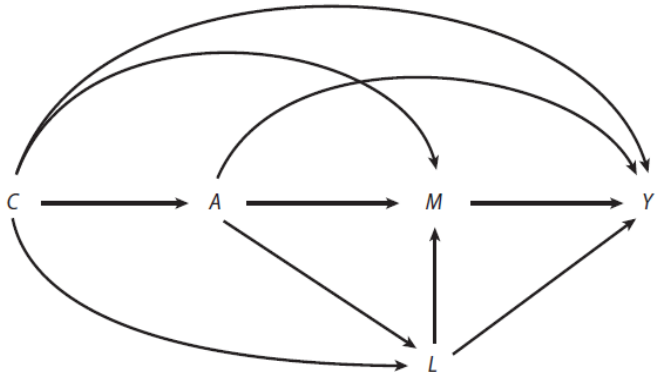
- Note, that this is different from applying the approach to mediation for a single mediator described earlier when dealing with one mediator at a time and then summing up the indirect effects
- This is because if the mediators were handled one at a time, then a different regression for Y would be fit for each mediator and only one mediator would be included in each of these regressions
  - These two approaches will coincide **if the mediators do not affect one another** (or more precisely, if the mediators are independent of one another conditional on A and C) but they will diverge otherwise:
  - See VanderWeele TJ & Vansteelandt S, 2013. Mediation analysis with multiple mediators. Epidemiol. Methods 2:95–115

INVERSE PROBABILITY-WEIGHTED FOR MEDIATION ANALYSES

- Instead of using traditional multivariable regression approaches, it is possible to fit Marginal Structural Models (MSMs) for decomposing the Total Effect.
- First, we estimate the exposure A weight:
  - $w_i^A = \frac{P(A=a_i)}{(p(A=a_i|C=c_i)}$
- Second, we estimate the mediator M weight:
  - $w_i^M = \frac{P(M=m_i \mid A=a_i)}{(p(M=m_i|A=a_i,C=c_i)}$
- Can be applied to multiple mediators
  - See: VanderWeele TJ & Vansteelandt S, 2013. Mediation analysis with multiple mediators. Epidemiol. Methods 2:95–115

MARGINAL STRUCTURAL MODELS FOR CDE IN THE PRESENCE OF EXPOSURE INDUCED CONFOUNDING (ASSUMPTION A.4)

- Fitting a MSM for the CDE
- $w_i^A = \frac{P(A=a_i)}{(p(A=a_i|C=c_i)}$
- $w_i^M = \frac{P(M=m_i \mid A=a_i)}{(p(M=m_i|A=a_i,C=c_i,L=l_i)}$
- $w_i = w_i^A \times w_i^M$



- We can then employ the MSM approach to CDE by fitting regression model of Y on A and M using as w_i weights
- For NIE, we can then apply a product method using MSM
- Effect decomposition with exposure induced confounding
  - Different potential approaches: Joint Mediators; Interventional Effects; Weighting approaches
  - See: *VanderWeele et al. 2014.* Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder

INVERSE ODDS RATIO WEIGHTING FOR MULTIPLE MEDIATORS

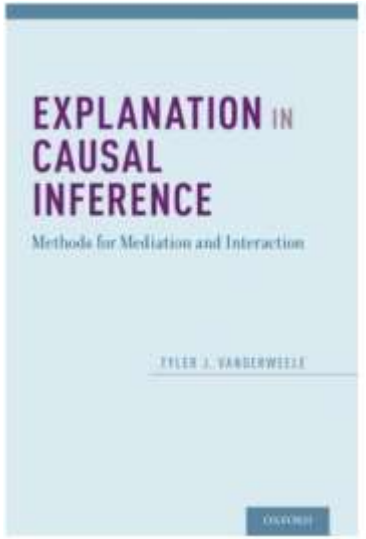
- An alternative approach that capitalizes on some properties of the OR
- Natural direct effect: outcome model on exposure with weight equal to the inverse of the covariate-adjusted exposure-mediator odds ratio association–indirect pathways of the mediators deactivated
  - Inverse odds ratio weight is calculated by regressing exposure on mediators and confounders, then taking the inverse of the predicted odds ratio
  - We can also get Stabilized inverse odds ratio weight (inverse odds weight)
- Total effect: outcome model on exposure and pre-exposure confounders of exposure and outcome
- Natural indirect effect: total effect – natural direct effect
- Benefits:
  - Given the invariance property of odds ratio, we could make the exposure and multiple mediators independent with weights estimated with one exposure-mediator model—easy when multiple mediators of different nature (e.g., categorical, continuous, binary) exist
  - Agnostic towards interactions between any mediator and the exposure on the outcome
- Disadvantages:
  - Only for binary exposures
  - Variances of estimates can be wider than those of traditional methods

Tchetgen Tchetgen EJ. 2013. doi:10.1002/sim.5864. Nguyen QC. doi:10.1093/aje/kwu278.

INTERACTION AND MEDIATION

INCLUDING AN INTERACTION BETWEEN THE EXPOSURE AND THE MEDIATOR

- This is possible to include an interaction between the exposure and the mediator
- The formulas need to be modified to include the interaction term but the interpretation is similar
  - See p.45-46 Explanation in Causal Inference Book
- In addition, the total effect can be decomposed:
  - Three way decomposition (VanderWeele, 2013)
  - Four Way decomposition (VanderWeele, 2014)



VanderWeele, Tyler. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.

THREE-WAY DECOMPOSITION

(VANDERWEELE, EPIDEMIOLOGY 2013, 224-31)

E binary (0,1); M binary (0;1)

$$(Y_1 - Y_0) = (Y_{1M_0} - Y_{0M_0}) + (Y_{0M_1} - Y_{0M_0}) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1 - M_0)$$

- $(Y_{1M_0} - Y_{0M_0})$ is the pure direct effect
- $(Y_{0M_1} - Y_{0M_0})$ is the pure indirect effect
- $(Y_{11} - Y_{10} - Y_{01} + Y_{00})$ is the counterfactual measure of the additive interaction between the exposure and the mediator
- $(M_1 - M_0)$ is the effect of the exposure on the mediator

FOUR-WAY DECOMPOSITION

VANDERWEELE. EPIDEMIOLOGY 2014; 25: 749-61

E binary (0,1); M binary (0;1)

$$(Y_1 - Y_0) = (Y_{10} - Y_{00}) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_0) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1) + (Y_{01} - Y_{00})(M_1 - M_0)$$

→ Proportion of exposed to M=0

→ Proportion of exposed to M=1

→ [note: $(Y_{01} - Y_{00})(M_1 - M_0) = (Y_{0M_1} - Y_{0M_0})$]

- The 1st term is the CDE when M is fixed to 0
- The 2nd term is the reference interaction, i.e. the additive interaction that operates when M is naturally present (i.e. the mediator is present in absence of the exposure)
- The 3rd term is the mediated interaction, i.e the additive interaction that operates through the effect of the exposure on the mediator
- The 4th term is the pure indirect effect

$$\mathbf{TE = CDE + INT_{ref} + INT_{med} + PIE}$$

PRACTICAL EXAMPLE

E	M	Risk of Y	Total population
0	0	0.01	100
1	0	0.03	100
0	1	0.04	100
1	1	0.08	100

$$(Y_1 - Y_0) = (Y_{10} - Y_{00}) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_0) + (Y_{11} - Y_{10} - Y_{01} + Y_{00})(M_1) + (Y_{01} - Y_{00})(M_1 - M_0)$$

[note: $(Y_{01} - Y_{00})(M_1 - M_0) = (Y_0M_1 - Y_0M_0)$]

CDE = 0.03 – 0.01 = 0.02
Ref. Interaction = (0.08-0.04-0.03+0.01) * M_0 = 0.01
Mediated interaction = (0.08-0.04-0.03+0.01) * (M_1) = 0.01
Pure indirect effect = 0.04-0.01 = 0.03
TE = 0.07

RELATION TO STATISTICAL MODELS

- Suppose that *Y* and *M* are continuous, and that the following regression models for *Y* and *M* are correctly specified:

$$E[Y \mid a, m, c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$
$$E[M \mid a, c] = \beta_0 + \beta_1 a + \beta_2' c.$$

- for exposure levels *a* and *a**, and setting the mediator to 0 in the controlled direct effect, the 4 components are given by:

$$E[CDE \mid c] = \theta_1 (a - a^*)$$
$$E[INT_{ref} \mid c] = \theta_3 (\beta_0 + \beta_1 a^* + \beta_2' c)(a - a^*)$$
$$E[INT_{med} \mid c] = \theta_3 \beta_1 (a - a^*)(a - a^*)$$
$$E[PIE \mid c] = (\theta_2 \beta_1 + \theta_3 \beta_1 a^*)(a - a^*).$$

If the exposure is binary (0,1), we have:

CDE = θ_1

Ref. Interaction = θ_3 ($\beta_0 + \beta'_2 c$)

Mediated interaction = $\theta_3 \times \beta_1$

Pure indirect effect = $\theta_2 \times \beta_1$

STATA CODE FOR THE 4-WAY DECOMPOSITION

```
/*************************************************************/
// 4-Way Decomposition Program
//
// BINARY EXPOSURE, CONTINUOUS MEDIATOR, CONTINUOUS OUTCOME
// Variables (parameters) sent to program:
// 1 = Exposure          -- here Binary, a = 1 & a* = 0
// 2 = Mediator -- continuous [ you would need to change the regression models)
// 3 = Outcome  -- continuous
// 4 through 6 = covariates C1-C3 (4=C1; 5=C2; 6=C3) you can add more as needed by including them in the different regressions)
// 7 = mstar is a scalar value at which mediator will be controlled at, can choose center at a value of 0 or choose the median as we did..
//
/*************************************************************/
  capture program drop decomp_cont_cont

  program decomp_cont_cont, rclass
      capture drop aXm
      gen aXm = `1' * `2'
  // outcome model                E[Y|a,m,c]
      regress `3' `1' `2' aXm `4' `5' `6'
      scalar AonY = _b[`1']
      scalar MonY = _b[`2']
      scalar IonY = _b[aXm]

      return scalar cde = AonY + IonY*`7'            // Controlled Direct Effect     -- E[CDE(m*)|C]

      capture drop samp                              // ensures mediator model estimated on same sample
      drop if e(sample) == 0                         // can be removed with complete data

        // mediation model             E[M|a,c]
          regress `2' `1' `4' `5' `6'
          scalar intercept = _b[_cons]
          scalar b_c1 = _b[`4']
          scalar b_c2 = _b[`5']
          scalar b_c3 = _b[`6']
          scalar AonM = _b[`1']

          return scalar pie = AonM*MonY                  // Pure Indirect Effect     -- E[PIE|c] NOTE: IonY drops out if a* = 0
          return scalar med_int = AonM*IonY              // Mediated Interaction Effect  -- E[INT_med|c]

          summarize `4'
          scalar m_c1 = r(mean)
          summarize `5'
          scalar m_c2 = r(mean)
          summarize `6'
          scalar m_c3 = r(mean)

          return scalar pde = AonY + IonY*(intercept + b_c1*m_c1 + b_c2*m_c2 + b_c3*m_c3)          // Pure Direct Effect -- E[PDE|c] NOTE: EonM drops out if a* = 0
          return scalar ref_int = IonY * (intercept + b_c1*m_c1 + b_c2*m_c2 + b_c3*m_c3 - `7')     // Reference Interaction Effect -- E[INT_ref|c]
          // return scalar ref_int = pde - (EonY + IonY*`7') -- alternate code
      end
```

```
To call the program and obtain 95% CIs, :
bootstrap CDE = r(cde), reps(500) seed(42782): decomp_cont_cont EXPOSURE MEDIATOR OUTCOME C1 C2 C3 MSTAR
bootstrap PIE = r(pie), reps(500) seed(42782): decomp_cont_cont EXPOSURE MEDIATOR OUTCOME C1 C2 C3 MSTAR
bootstrap MIE = r(med_int), reps(500) seed(42782): decomp_cont_cont EXPOSURE MEDIATOR OUTCOME C1 C2 C3 MSTAR
bootstrap RIE = r(ref_int), reps(500) seed(42782): decomp_cont_cont EXPOSURE MEDIATOR OUTCOME C1 C2 C3 MSTAR
```

R CODE FOR THE 4-WAY DECOMPOSITION

```
#-----

# All libraries here
library(   boot )
library(   survival   )
library(   data.table   )
library(   foreign   )
library(   dummies   )
library(   GenABEL   )
library(   dummies   )
#-------------------------------------------------------------------------------------------------------------
------------
# Sources import here

# this script should be run from the same folder where src.R is
source(   'src .R ' )
#-------------------------------------------------------------------------------------------------------------
------------
# Define your parameters here!!!

#Data pathway
data_path    <- "//storage.erasmusmc.nl/m/MyDocs/592004/My Documents/Desktop/Rscript4way/Test.sav                    "

#Path to save results
output   <- '//storage.erasmusmc.nl/m/MyDocs/592004/My Documents/Desktop/Rscript4way/Test_results.csv                    '

#Define variables
A << - 'A2 '
M << - 'M1 '
Y << - 'Y1 '
COVAR   << - c( 'C1 ', 'C2 ', 'C3 ')



#1=binary 0=continuous
outcome   =0
mediator   =0

#Assign levels for the exposure that are being compared;
#for mstar it is the level at which to compute the CDE and the remainder of the decomposition
a << - 1
astar   << - 0
mstar   << - 0

#Boostrap number of iterations
N_r =5


####### DONT TOUCH FROM HERE
#######


# Reading data file
data   <- read.spss(    data_path    , to.data.frame     =T )  #TODO spss/csv/txt (?)


if   (!  prod(c(   A , Y , M , COVAR   ) %in%    names(   data   ) ) )  {stop(     'Some of defined variable names are not in data
file!   ')}

if   (   mediator    == 1   &   outcome    == 1   ) {   save_results(       output   =output   ,   boot_function     =boot.bMbO    ,   N =N_r )   }
if   (   mediator    == 0   &   outcome    == 1   ) {   save_results(       output   =output   ,   boot_function     =boot.cMbO    ,   N =N_r )   }
if   (   mediator    == 1   &   outcome    == 0   ) {   save_results(       output   =output   ,   boot_function     =boot.bMcO    ,   N =N_r )   }
if   (   mediator    == 0   &   outcome    == 0   ) {   save_results(       output   =output   ,   boot_function     =boot.cMcO    ,   N =N_r )   }
```

https://github.com/Unalmut/4way-decomposition

SAS CODE FOR THE 4-WAY DECOMPOSITION

Continuous Outcome, Continuous Mediator

```
proc nlmixed data=mydata;
parms t0=0 t1=0 t2=0 t3=0 tc1=0 tc2=0 tc3=0 b0=0 b1=0 bc1=0 bc2=0 bc3=0 ss_m=1 ss_y=1;
a1=1; a0=0; mstar=0; cc1=10; cc2=10; cc3=20;
mu_y=t0 + t1*A + t2*M + t3*A*M + tc1*C1 + tc2*C2 + tc3*C3;
mu_m =b0 + b1*A + bc1*C1 + bc2*C2 + bc3*C3;
ll_y= -((y-mu_y)**2)/(2*ss_y)-0.5*log(ss_y);
ll_m= -((m-mu_m)**2)/(2*ss_m)-0.5*log(ss_m);
ll_o= ll_m + ll_y;
model Y ~general(ll_o);
bcc = bc1*cc1 + bc2*cc2 + bc3*cc3;
cde = (t1 + t3*mstar)*(a1-a0);
intref = t3*(b0 + b1*a0 + bcc - mstar)*(a1-a0);
intmed = t3*b1*(a1-a0)*(a1-a0);
pie = (t2*b1 + t3*b1*a0)*(a1-a0);
te = cde + intref + intmed + pie;
estimate 'Total Effect' te;
estimate 'CDE' cde;
estimate 'INTref' intref;
estimate 'INTmed' intmed;

estimate 'Proportion CDE' cde/te;
estimate 'Proportion INTref' intref/te;
estimate 'Proportion INTmed' intmed/te;
estimate 'Proportion PIE' pie/te;
estimate 'Overall Proportion Mediated' (pie+intmed)/te;
estimate 'Overall Proportion Attributable to Interaction' (intref+intmed)/te;
estimate 'Overall Proportion Eliminated' (intref+intmed+pie)/te;
run;
```

MEDIATION ANALYSIS FOR ENVIRONMENTAL JUSTICE STUDIES

MEDIATION ANALYSES TO DECOMPOSE HEALTH INEQUALITIES

Multiple mediators approach to study environmental chemicals as determinants of health disparities

Andrea Bellavia[a,b], Ami R. Zota[c], Linda Valeri[d,e], Tamarra James-Todd[a,f,g*]

American Journal of Epidemiology

Practice of Epidemiology

Mediation Analysis for Health Disparities Research

Ashley I. Naimi*, Mireille E. Schnitzer, Erica E. M. Moodie, and Lisa M. Bodnar

Socioeconomic and Tobacco Mediation of Ethnic Inequalities in Mortality over Time

Repeated Census-mortality Cohort Studies, 1981 to 2011

Tony Blakely,[a] George Disney,[a] Linda Valeri,[b] June Atkinson,[a] Andrea Teng,[a] Nick Wilson,[a] and Lyle Gurrin[c]

Inferential challenges when assessing racial/ethnic health disparities in environmental research

Tarik Benmarhnia[1*], Anjum Hajat[2] and Jay S. Kaufman[3]

WHY DECOMPOSING HEALTH DISPARITIES?

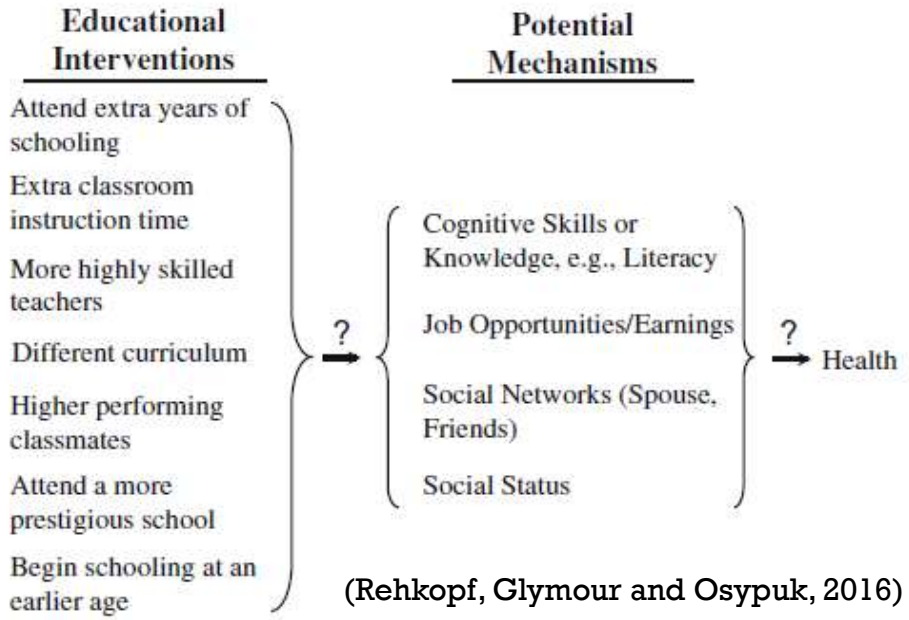
- Beyond highlighting and quantifying disparities in health
- We need to understand the determinants of disparities in health to orient specific interventions
- **The determinants of population health are not the same as the determinants of disparities in health**

"NON-MANIPULABLE" EXPOSURES, THE EXAMPLE OF RACE/ETHNICITY

- Reference to race/ethnicity in epidemiological research has grown since the 70's in the US (Jones et al. 1991)
  - About how to define R/E categories (Stolley 1999) or treat them in analytical models (Kaufman & Cooper 1999)
- R/E as an exposure can be seen as ambiguous
- Estimating the effect of R/E seems more conceivable if R/E represents the <u>experience of individual discrimination</u> or contextual segregation
  - Example of self-reported race for job applications (Bertrand and Mullainathan 2004)

SES VARIABLES AND THE CONSISTENCY ASSUMPTION

- Education, Income, neighborhood level of income etc..
- Violating the consistency assumption
- The presence of multiple versions of treatment is problematic:
  1. If the causal effect varies across versions: the magnitude of the average causal effect (ACE) depends on the proportion of individuals who received each version
  2. For the interpretation of the causal effect of interest



(Rehkopf, Glymour and Osypuk, 2016)

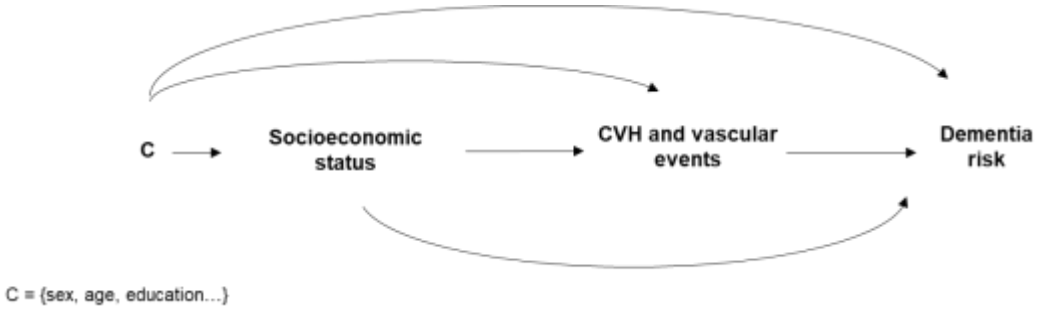
MEDIATION AS A SOLUTION AND IMPLICATIONS FOR DECOMPOSITION TECHNIQUES

- In this context, it becomes difficult to formally define direct and indirect effects in terms of potential outcomes
- VanderWeele and Robinson (2014) proposed to:
  - Estimate direct and mediated inequality measures
  - If we have the following linear model: $E[Y \mid a,m,c] = \theta_1 a + \theta_2 m + \theta_3' c$
  - *a* would be a binary R/E variable [NH blacks or white], *m* would be an indicator of adulthood SES
  - We can interpret θ_1 as the R/E inequality in Y remaining had the distribution of SES for the NH black population been set to equal that of the NH white population with the same values of C.
  - We can also estimate the magnitude of the **mediated inequality effect** through SES by comparing the inequality measure before/after controlling for SES and mediator-outcomes confounders

EXAMPLE: DETERMINANTS OF DEMENTIA INEQUALITIES

Socioeconomic inequalities in dementia risk among a French population-based cohort: quantifying the role of cardiovascular health and vascular events

Noémie Letellier[1] · Sindana D. Ilango[2,3] · Marion Mortamais[1] · Christophe Tzourio[4] · Audrey Gabelle[1,6] · Jean-Philippe Empana[5] · Cécilia Samieri[4] · Claudine Berr[1,6] · Tarik Benmarhnia[3,7]



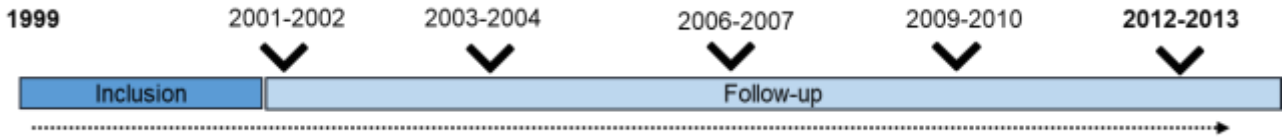


Table S1: Differences between the main cohort and the final study population (N=7324)

	Excluded (n = 1650)	Included (n = 5734)
Individual characteristics, n (%)		
Age (years)[a]	76 (71-80)	73 (69-77)
Women	842 (51.0)	3604 (62.9)
Study center		
Bordeaux	488 (29.6)	1088 (19.0)
Dijon	607 (36.8)	3124 (55.8)
Montpellier	555 (33.6)	1448 (25.3)
Primary education	537 (32.6)	1324 (23.1)
Blue-collar workers	393 (23.8)	943 (16.5)
Lower income	1122 (68.0)	3665 (63.9)
APOEƐ4 carrier	345 (20.9)	1121 (19.6)
Neighborhood characteristics*		
Unemployment rate	14.1 (10.3-19.1)	12.1 (9.6-16.6)
Proportion of blue-collar workers	15.8 (11.6-22.7)	15.6 (11.7-23.3)
Proportion of households without a car	25.1 (16.8-31.7)	24.3 (17.0-30.9)
Proportion of single parents	15.0 (11.9-19.5)	14.2 (11.2-18.2)
Proportion of tenants	52.8 (38.7-65.7)	51.2 (38.3-62.2)
Proportion of resident without secondary education	46.9 (39.0-59.6)	46.3 (39.3-59.5)

Table 2 Total effects of socioeconomic level on dementia risk (Weighted Cox proportional hazards model and Aalen model[a], N = 5581)

	All-type incident dementia (n = 515) n, (%)	Cox PH model		Aalen model	
		HR	(95% CI)	Estimate	(95% CI)
Educational level					
Secondary or higher	352/4309 (8.2)	ref	–		
Primary	163/1272 (12.8)	1.60	(1.44–1.78)	571×10^{-5}	$(269 \times 10^{-5}–873 \times 10^{-5})$
Occupational category					
White-collar	401/4670 (8.6)	ref	–		
Blue-collar	114/911 (12.5)	1.62	(1.43–1.84)	634×10^{-5}	$(246 \times 10^{-5}–1020 \times 10^{-5})$
Income (n = 5570)					
Higher income	299/3647 (8.2)	ref	–		
Lower income	216/1934 (11.2)	1.23	(1.09–1.39)	221×10^{-5}	$(-79 \times 10^{-5}–521 \times 10^{-5})$

HR hazard ratio; PH: Proportional Hazard; *CI* confidence interval

[a]Adjusted for age, gender, education (for income and occupation models), living alone (for income models), APOE4 and study centers

Table 3 Total, direct and indirect effects of socioeconomic level through CVH score in binary and each component (using IORW[a], N = 5581)

	Total effect		Natural direct effect		Natural indirect effect	
	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)
Primary education						
CVH score[b]	1.62	[1.32–1.99]	1.51	[1.22–1.88]	1.08	[1.03–1.14]
Components of CVH score						
Biological components	1.62	[1.32–1.99]	1.61	[1.31–1.99]	1.01	[1.00–1.02]
Behavioral components	1.62	[1.32–1.99]	1.60	[1.30–1.98]	1.01	[0.99–1.04]
Biological components						
Total cholesterol	1.62	[1.32–1.99]	1.62	[1.32–1.99]	1.00	[0.99–1.01]
Blood pressure	1.62	[1.32–1.99]	1.60	[1.30–1.95]	1.01	[1.00–1.03]
Fasting plasma glucose	1.62	[1.32–1.99]	1.61	[1.31–1.98]	1.01	[1.00–1.02]
Behavioral components						
Smoking	1.62	[1.32–1.99]	1.66	[1.35–2.03]	0.98	[0.94–1.00]
Healthy diet	1.62	[1.32–1.99]	1.56	[1.26–1.94]	1.04	[0.99–1.09]
Physical activity	1.62	[1.32–1.99]	1.63	[1.33–1.99]	1.00	[0.98–1.01]
BMI	1.62	[1.32–1.99]	1.61	[1.31–2.00]	1.01	[0.97–1.05]

WHAT ABOUT ENVIRONMENTAL JUSTICE STUDIES?

- When understanding the drivers of health inequalities across 2 or multiple groups, we can focus on environmental drivers of such inequalities
- We can consider the role of differential exposure and differential susceptibility
  - The Double Jeopardy of Environmental Justice
- In this setting, we can implement decomposition techniques we discussed today
  - For one or multiple mediators
  - Including the exposure-mediator interaction
    - To handle differential susceptibility

Differential vulnerability and susceptibility: how to make use of recent development in our understanding of mediation and interaction to tackle health inequalities

Finn Diderichsen,[1,2]* Johan Hallqvist[3] and Margaret Whitehead[4]

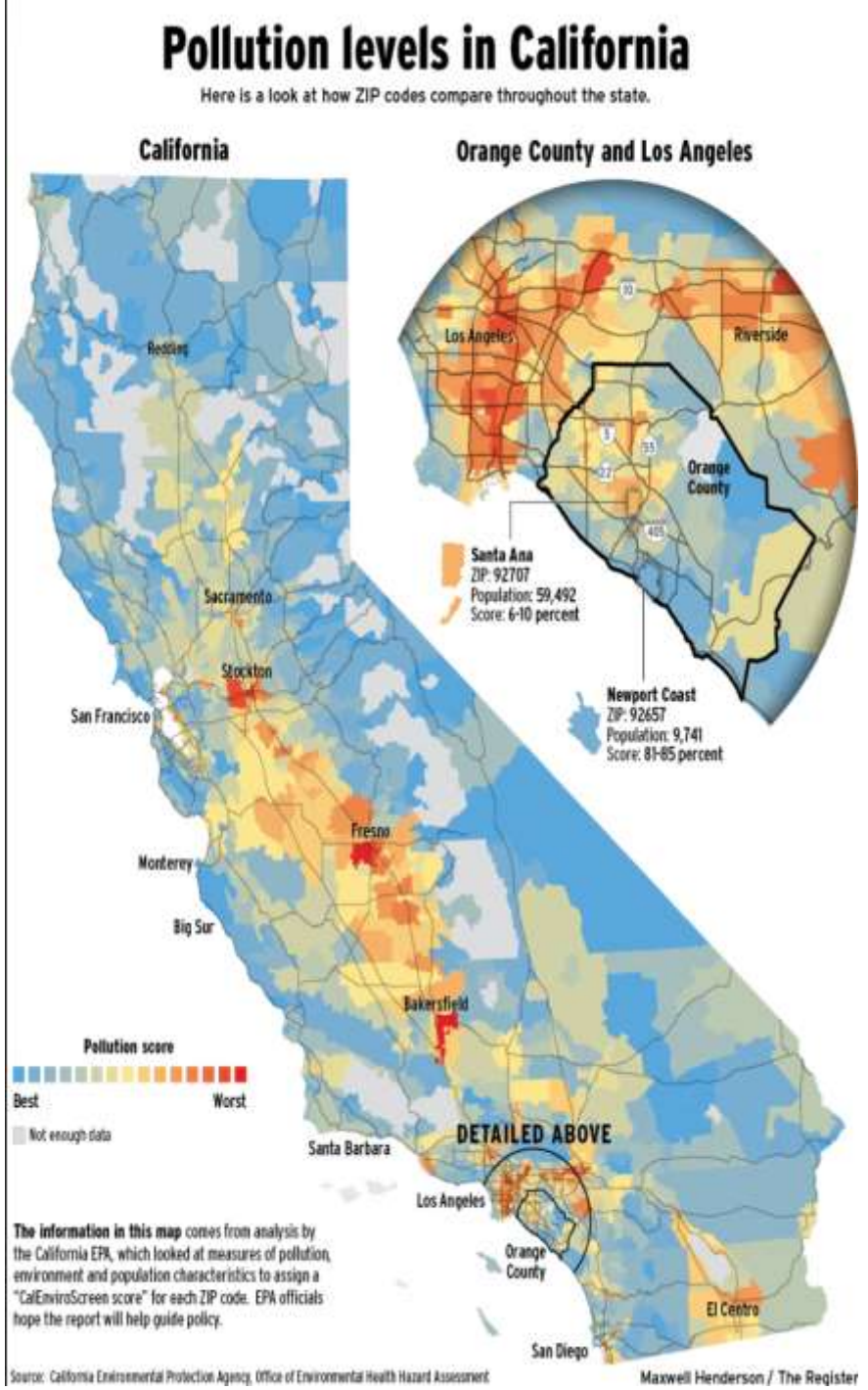
CASE STUDY #2:

DECOMPOSITION ANALYSIS OF BLACK–WHITE DISPARITIES IN BIRTH OUTCOMES: THE RELATIVE CONTRIBUTION OF AIR POLLUTION AND SOCIAL FACTORS IN CALIFORNIA

Benmarhnia, T., Huang, J., Basu, R., Wu, J., & Bruckner, T. A. (2017). Decomposition analysis of Black–White disparities in birth outcomes: the relative contribution of air pollution and social factors in California. Environmental health perspectives, 125(10), 107003.

DATA OVERVIEW

- Data from the California birth certificates (2005-2010)
- Birth outcomes.
  - Preterm Birth (PTB): delivered <37 weeks of gestation
  - very-preterm birth (VPTB): delivery between 20 and 33 completed weeks of gestation
  - Small for gestational age (SGA) = sex-specific, birth weight < 10th percentile for the given gestational age
- Other individual characteristics
  - Maternal age, level of education, Medicaid insurance status, single parenthood …
- Linked to census variables for neighborhood characteristics
- And to air pollution modelled at the Zip code level
  - Data collected for: PM2.5 (in µg/m3); NO2 (in ppm).



MAIN RESULTS

Table 1. Study population characteristics of all singleton live births in California from 2005–2010, by maternal non-Hispanic black or white status [*n*(%)].

Characteristic	Non-Hispanic black	Non-Hispanic white
Birth outcomes		
Preterm birth		
No	110 247 (87.5)	577354 (89.8)
Yes	20 423 (12.5)	65589 (10.2)
Very preterm birth		
No	148585 (97.4)	772079 (98.9)
Yes	3914 (2.5)	8434 (1.1)
Small for gestational age		
No	143275 (81.7)	803412 (90.1)
Yes	32022 (18.3)	88074 (9.9)

Table 3. Predicted probability and disparity in PTB, VPTB, and SGA between non-Hispanic black and non-Hispanic white mothers, and percentage of the difference explained by individual, neighborhood socioeconomic, and neighborhood air pollution variables (California 2005–2010).

Race/ethnicity or explanatory variable	PTB estimate (95% CI)	VPTB estimate (95% CI)	SGA estimate (95% CI)
Predicted probability			
Non-Hispanic black mother	0.157 (0.153, 0.159)	0.026 (0.025, 0.027)	0.183 (0.180, 0.185)
Non-Hispanic white mother	0.101 (0.101, 0.102)	0.011 (0.010, 0.011)	0.099 (0.098, 0.100)
Black-white disparity	0.056 (0.054, 0.058)	0.015 (0.014, 0.016)	0.084 (0.081, 0.087)
Percent difference explained			
Total difference explained	39.3%	31.1%	30.8%
Maternal socioeconomic predictors	17.5%	10.5%	30.8%
Educational attainment	4.8 (3.7, 5.7)	4.0 (2.7, 5.6)	4.0 (3.6, 4.4)
Age at delivery	4.6 (3.4, 5.9)	0.3 (−1.1, 2.0)	3.1 (2.7, 3.5)
Medicaid enrollee	5.4 (4.0, 6.7)	1.7 (0.4, 2.9)	5.2 (4.7, 5.7)
Missing paternal information	2.7 (2.1, 3.3)	4.4 (3.2, 5.8)	2.0 (1.7, 2.3)
Neighborhood socioeconomic environment predictors	16.1%	13.2%	12.1%
Unemployment rate	3.8 (2.2, 5.3)	3.4 (1.5, 5.0)	3.2 (2.5, 3.9)
Poverty rate	4.8 (3.1, 6.4)	3.9 (1.9, 5.6)	3.8 (3.0, 4.5)
Linguistic minority	1.9 (0.3, 3.4)	2.0 (0.3, 3.5)	1.4 (0.7, 2.1)
Educational attainment	5.5 (4.5, 6.4)	3.9 (1.9, 5.7)	3.7 (2.9, 4.5)
Neighborhood air pollution predictors	5.7%	7.4%	4.3%
NO_2 concentration	2.6 (−2.1, 5.1)	2.9 (−1.3, 3.9)	0.3 (−0.4, 1.2)
$PM_{2.5}$ concentration	3.1 (0.9, 5.2)	4.5 (0.7, 7.8)	4.0 (1.0, 6.8)

PART II - MEDIATION ANALYSIS IN PRACTICE

BREAK

PART III
SUMMARY AND ADVANCED TOPICS

A CAUTIONARY NOTE ON EXISTING PACKAGES

- Several packages are available
- In R:
  - CMAverse: six causal mediation analysis (https://bs1125.github.io/CMAverse/)
  - mediation (Tingley et al.) https://cran.r-project.org/web/packages/mediation/index.html
- Paramed in Stata
  - Emsley R, Liu H. PARAMED: Stata module to perform causal mediation analysis using parametric regression models. Stat Softw Components 2013
- Causalmed in SAS
  - Valeri, L. and VanderWeele, T.J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18:137-150

NEW GUIDELINES FOR MEDIATION ANALYSES

JAMA | Special Communication

A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies
The AGReMA Statement

Hopin Lee, PhD; Aidan G. Cashin, PhD; Sarah E. Lamb, DPhil; Sally Hopewell, DPhil; Stijn Vansteelandt, PhD; Tyler J. VanderWeele, PhD; David P. MacKinnon, PhD; Gemma Mansell, PhD; Gary S. Collins, PhD; Robert M. Golub, MD; James H. McAuley, PhD; and the AGReMA group

Key Points

Question What information should be reported in studies that include mediation analyses of randomized trials and observational studies?

Findings An international Delphi and consensus process (using the Enhancing Quality and Transparency of Health Research methodological framework) generated a 25-item reporting guideline for primary reports of mediation analyses and a 9-item short form for secondary reports of mediation analyses.

Meaning Using the 25-item or 9-item reporting guideline may facilitate peer review and could help ensure that studies using mediation analyses are completely, accurately, and transparently reported.

SETTINGS NOT DISCUSSED TODAY

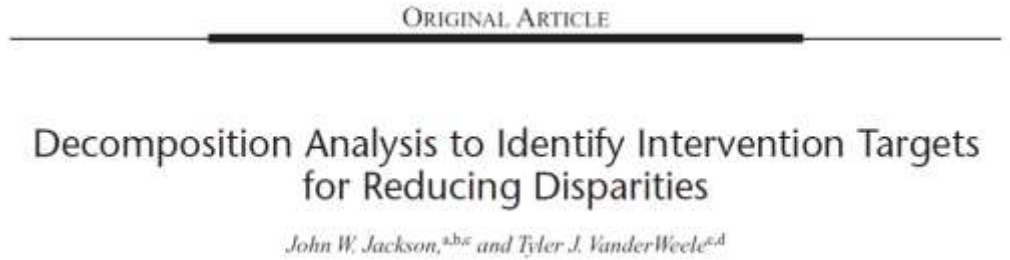
- Approaches we discussed today could be extended to various settings:
  - Non-linear settings
  - Mediator-Mediator Interactions ..
  - Time to event models
    - Aalen et al. (2020). Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. Biometrical Journal, 62(3), 532-549.
- High-dimensional settings
  - Omics data [would deserve an entire workshop by itself]
  - Mediation analyses coupled with mixtures methods
- Sensitivity analyses for mediation analyses
  - See additional slides

ALTERNATIVE APPROACHES FOR MEDIATION ANALYSES

- Alternative weighting approaches
  - Lange, Theis, Stijn Vansteelandt, and Maarten Bekaert. "A simple unified approach for estimating natural direct and indirect effects." American journal of epidemiology 176.3 (2012): 190-195.
  - Inverse odds ratio weighting: Nguyen, Quynh C., et al. "Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting." American journal of epidemiology 181.5 (2015): 349-356.
- Standardization approaches (e.g. G-computation)
  - Wang, Aolin, and Onyebuchi A. Arah. "G-computation demonstration in causal mediation analysis." *European journal of epidemiology* 30.10 (2015): 1119-1127.
- Stochastic mediation approaches
  - Rudolph, Kara E., et al. "Robust and flexible estimation of stochastic mediation effects: a proposed method and example in a randomized trial setting." *Epidemiologic Methods* 7.1 (2017).
  - A very active area of research

OTHER TYPES OF DECOMPOSITION

- Alternative approaches exist in other disciplines like demography or econometrics
- Oaxaca and Blinder (1973) developed a regression based decomposition to partition the gap in an outcome of interest between two groups into an "explained" and an "unexplained" portion
  - The "explained" portion of the gap is the difference in the outcome attributable to group differences in levels of a set of measured predictor variables between the "advantaged" and the "disadvantaged" groups
  - The "unexplained" portion arises from differentials in how the predictor variables are associated with the outcomes for the two groups.
- Mediation analyses and Oaxaca-Blinder decomposition can be equivalent under certain conditions


ORIGINAL ARTICLE

Decomposition Analysis to Identify Intervention Targets for Reducing Disparities

John W. Jackson,[a,b,c] and Tyler J. VanderWeele[c,d]

SUMMARY FOR TODAY

THANKS

CHC048@UCSD.EDU
TBENMARHNIA@UCSD.EDU

ευχαριστώ

ADDITIONAL SLIDES

A SIDE NOTE ON TREATMENT CONFOUNDER FEEDBACK AND INVERSE PROBABILITY OF TREATMENT WEIGHTING (IPTW)

WHEN TRADITIONAL METHODS TO ADJUST FOR CONFOUNDING DO NOT WORK

- There is time-varying confounding for the causal effects of treatment strategy $\bar{a}$ if:

$$E[Y^{\bar{a}}] \neq E[Y|A = \bar{a}]$$

- i.e., if the mean outcome had all individuals in the study followed strategy $\bar{a}$ (i.e., the counterfactual) differs from the mean outcome among the subset of individuals who followed strategy $\bar{a}$ in the actual study (i.e., observed).

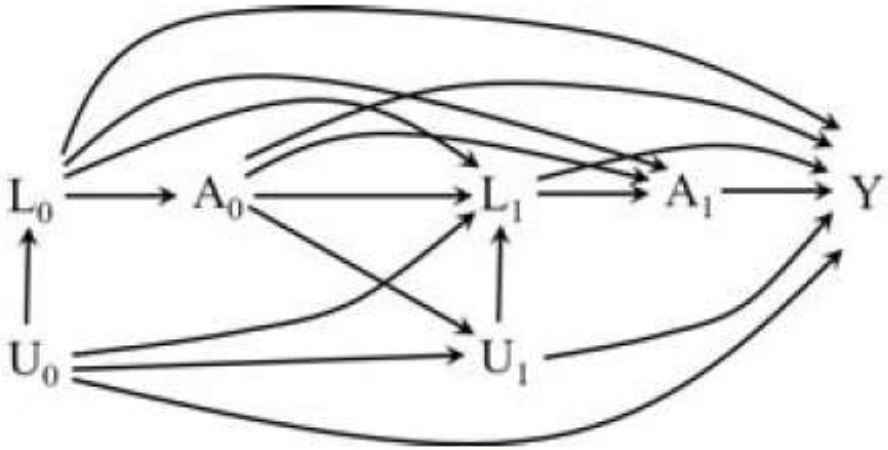


Figure 20.1

SEQUENTIAL EXCHANGEABILITY

- Causal inference with time-varying treatments requires adjusting for the time-varying covariates L_k to achieve conditional exchangeability at each time point k.
- In a study with two time points, sequential conditional exchangeability is the combination of conditional exchangeability at both the first time and the second time of the study.
- Formal definition of Sequential exchangeability for Y^g, i.e., the potential outcome under treatment strategy g:

$$Y^g \perp\!\!\!\perp A_k | A_{k-1} = g(\bar{A}_{k-2}, \bar{L}_{k-2}), \bar{L}_k \text{ for all strategies } g, \text{ and } k=0, 1, 2, \ldots K.$$

- This form of sequential exchangeability always holds in a *perfect* sequentially randomized trial.
- No form of sequential exchangeability is guaranteed to hold in observational studies, but it may be approximated if the probability of treatment A_k only depends on the treatment history and measured covariate history (see next slides).

TREATMENT CONFOUNDER FEEDBACK

- The confounder affects the treatment (A1), and the treatment (A0) affects the confounder.
- Suppose we want to estimate the causal effect of "always treat" vs. "never treat" strategy, i.e.,:

$$E[Y^{a_0=1,a_1=1}] - E[Y^{a_0=0,a_1=0}]$$

In DAG 20.3, a simplified version of DAG 20.1, there is no forward directed path from A to Y, i.e., the sharp null hypothesis is true.

However, adjustment for L1 would open a biasing pathway, that would result in a measured association between A and Y (i.e., bias under the null).

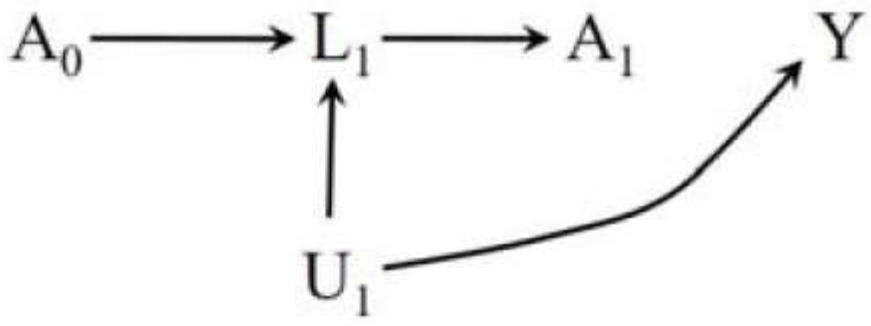


Figure 20.3

WHY TRADITIONAL METHODS FAIL

- When there are time-varying confounders and treatment-confounder feedback, **traditional methods cannot be used to correctly adjust for those confounders.**
- In the example on figure 20.5 with a hypothesized null effect, L_1 is a collider variable, so stratification or adjustment methods will result in **collider stratification bias**.
- In other words, confounding is *eliminated*, but at the same time, selection bias is *introduced*.

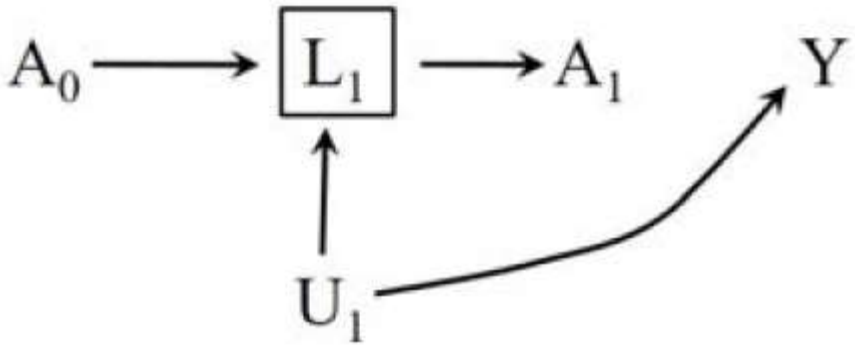


Figure 20.5

WHY TRADITIONAL METHODS FAIL

- Even when there is a non-null relationship, bias may be introduced.
- In DAG 20.1, for which there is a non-null effect:

1. Adjustment for L_1 opens a pathway from $A_0 \rightarrow L_1 \leftarrow U_1 \rightarrow Y$
2. Adjustment for L_1 prevents the estimation of the indirect effect $A_0 \rightarrow L_1 \rightarrow A_1 \rightarrow Y$

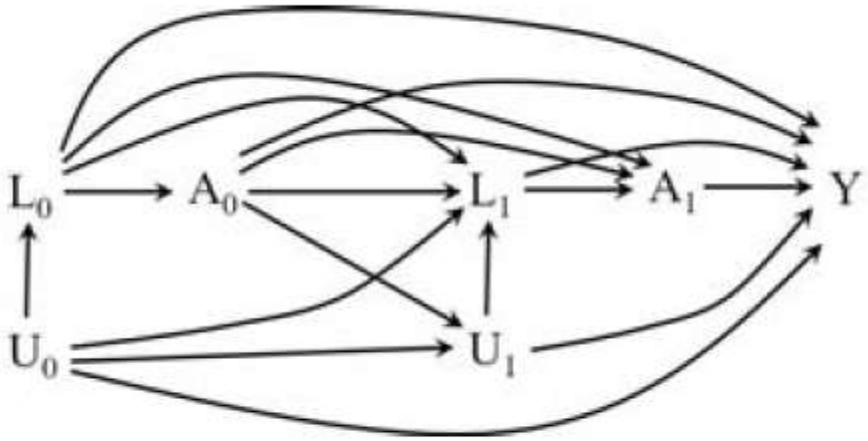


Figure 20.1

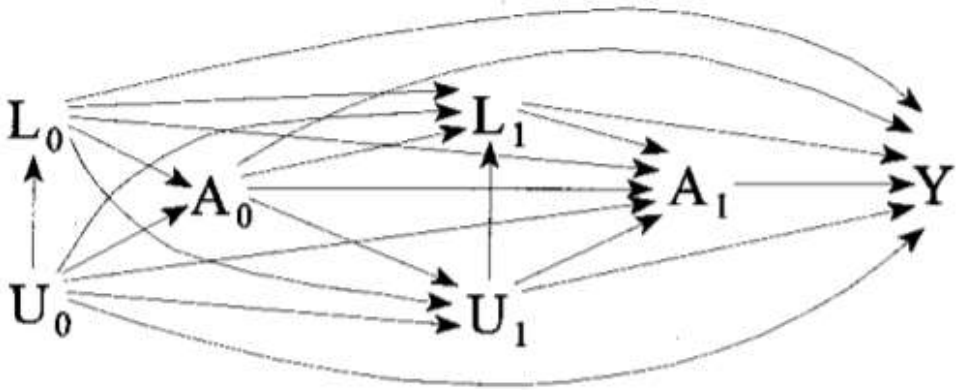
WHAT ARE G-METHODS?

- G-methods are designed for application to generalized treatment contrasts involving treatments that vary over time
  - Robins (1986) named such models "g methods" for "general", to enable the identification and estimation of the effects of all kinds of exposures or intervention plans.
- They typically include 3 types of methods:
  1. The G-formula or G-computation formula
     Non-parametric Standardization
     Parametric G-computation
  2. **Marginal Structural Models with IPW**
  3. G-estimation of Structural nested models

  +and their doubly-robust generalizations

WHEN IS IT USEFUL?

- Mostly: Time dependent confounding and exposures
- But also:
  - Censoring
  - Mediation
  - Simulating interventions with time-fixed or time-varying exposures



INVERSE PROBABILITY OF TREATMENT WEIGHTING

- The inverse probability of treatment weighting (IPTW) is estimated using the propensity score.
- Then, this weight can be used in the outcome model to consider differences at baseline regarding the (inverse) probability of receiving the treatment (i.e. the PS)
- A very useful tool in epidemiology in general
  - IPTW for time-fixed treatments and confounding
  - For MSM and time-varying treatments and confounding
  - Informed censoring

IPTW

- Weighting the outcome regression model by the inverse of the probability of treatment (IPTW) creates a **pseudo-population** in which the **arrow from the confounders to the treatment/exposure is removed**.
- Each subject is assigned a weight $W^A = {}^1/_{f(A|L)}$
- The result of the weighting procedure is a simulated pseudo-population in which all members of the study population are replaced by two copies of themselves – one who is treated and one who is untreated.
- The IP weights $W^A = {}^1/_{f(A|L)}$ adjust for confounding by L because they create a pseudo-population in which all individuals have the same probability of receiving A = 1 (a probability equal to 1) and the same probability of receiving A = 0 (also 1).

MODEL-BASED IPTW EXAMPLE

- NHEFS observational cohort study
- 1,566 smokers enrolled at baseline
- A = smoking cessation
- Y = weight gain
- Anyone who quit smoking during the follow-up period was classified as exposed (A=1); otherwise they were unexposed (A=0).
- Weight was measured at baseline and 10-year follow-up.

- **Average weight gain**
  - Quitters: $\hat{E}(Y|A=1) = 4.5$ kg
  - Non-quitters: $\hat{E}(Y|A=0) = 2.0$ kg
  - $\hat{E}(Y|A=1) - \hat{E}(Y|A=0) = 2.5$ kg
    - 95% CI: 1.7 to 3.4

MODEL-BASED IPTW EXAMPLE

- $E[Y^{a=1}]$ = mean weight gain that would have been observed if all individuals in the population had quit smoking
- $E[Y^{a=0}]$ = mean weight gain that would have been observed if no individuals in the population had quit smoking
- Average causal effect (risk difference):

$$E[Y^{a=1}] - E[Y^{a=0}]$$

- I.e., the difference in mean weight that would have been observed if everybody had been treated compared with untreated

MODEL-BASED IPTW EXAMPLE

Table 12.1

Mean baseline characteristics	A 1	0
Age, years	46.2	42.8
Men, %	54.6	46.6
White, %	91.1	85.4
University, %	15.4	9.9
Weight, kg	72.4	70.3
Cigarettes/day	18.6	21.2
Years smoking	26.0	24.1
Little exercise, %	40.7	37.9
Inactive life, %	11.2	8.9

- Do you expect the average difference in weight gain:

$$\hat{E}(Y|A=1) - \hat{E}(Y|A=0) = 2.5$$

- To be equivalent to the causal risk difference?

$$E[Y^{a=1}] - E[Y^{a=0}]$$

ESTIMATING THE IPTW

- *We assume that the following 9 variables, all measured at baseline, are sufficient to adjust for confounding:*
- *L1=sex (0: male, 1: female),*
- *L2=age (in years),*
- *L3=race (0: white, 1: other),*
- *L4=education (5 categories),*
- *L5=intensity of smoking (number of cigarettes per day),*
- *L6=duration of smoking (years of smoking),*
- *L7=physical activity in daily life (3 categories),*
- *L8=recreational exercise (3 categories), and*
- *L9=weight (in kg).*

ESTIMATING THE IPTW

1. Fit a Logistic regression model for the exposure on all covariates sufficient for confounding control
2. $\Pr(X = 1) = expit(\beta_0 + \beta_1 L1 + \beta_2 L2 + \beta_3 L2^2 + \beta_4 L3 + \beta_5 L4 + \cdots)$
2. Output the predicted probability (pred) from the logistic model, $\widehat{Pr}[A = 1|L]$ for each individual (N=1,566): **like a propensity score**
3. Set the weights: 1/pred for treated and 1/1-pred for the untreated
4. Fit a linear mean model for the outcome using each individual's IPW as the weight

$$E[Y|A] = \theta_0 + \theta_1 A$$

- In the example, θ_1 was estimated as 3.4 kg.
- 95% CIs can be obtained via:
  - Statistical theory
  - Bootstrapping
  - Robust variance estimation

With the third option, the 95% CI was estimated as: (2.4, 4.5)

STABILIZED AND TRUNCATED WEIGHTS

- IPW estimation can be unstable when some values of A are uncommon, then the weight denominator, $\Pr(A|L)$ can be very small.
- One solution is to use a stabilizing factor – this reduces the variance of the weights and results in more precise effect estimates.

UNSTABILIZED WEIGHTS

- Treated:
$$W^A = 1/\Pr(A = 1|L = l)$$
- Untreated:
$$W^A = 1/\Pr(A = 0|L = l)$$

Or

$$1/[1 - \Pr(A = 1|L = l)]$$

STABILIZED WEIGHTS

- Treated:
$$SW^A = \Pr(A = 1)/\Pr(A = 1|L = l)$$
- Untreated:
$$SW^A = \Pr(A = 0)/\Pr(A = 0|L = l)$$

Or

$$[1 - \Pr(\mathrm{A} = 1)]/[1 - \Pr(A = 1|L = l)]$$

- **Truncated Weights**
  - We can remove the extreme weights
  - Typically, we remove observations >99th and < 1st percentiles

THE OAXACA-BLINDER DECOMPOSITION

HOW DOES IT WORK? (1/3)

- Following Sen's (2014) example, we want to understand the BMI differences between NH blacks individuals and NH white individuals.
- We can model 2 separate linear models as follows:

$$\text{BMI}^{W}_{\text{mean}} = \beta_0^W + \sum_{j=1}^{J} .\beta_j^W X_j^W{}_{\text{mean}}$$

$$\text{BMI}^{B}_{\text{mean}} = \beta_0^B + \sum_{j=1}^{J} .\beta_j^B X_j^B{}_{\text{mean}}$$

- The Racial difference in BMI can be written as follows:

$$\text{BMI}^{B}_{\text{mean}} - \text{BMI}^{W}_{\text{mean}} = (\beta_0^B - \beta_0^W) + \sum_{j=1}^{J} (\beta_j^B X_j^B{}_{\text{mean}} - \beta_j^W X_j^W{}_{\text{mean}})$$

HOW DOES IT WORK? (2/3)

- The Oaxaca decomposition generates a hypothetical term with the mean X values of the NH whites, but the β of the NH blacks

$$BMI^{B}_{\text{mean}} - BMI^{W}_{\text{mean}} = [(\beta^{B}_{0} - \beta^{W}_{0}) + \sum_{j=1}^{J}(\beta^{B}_{j} X^{B}_{j\ \text{mean}} - \beta^{W}_{j} X^{W}_{j\ \text{mean}}) + \sum_{j=1}^{J} \beta^{B}_{j} X^{W}_{j\ \text{mean}} - \sum_{j=1}^{J} \beta^{B}_{j} X^{W}_{j\ \text{mean}}]$$

- Which can also be written as follows:

$$BMI^{B}_{\text{mean}} - BMI^{W}_{\text{mean}} = \left[\sum_{j=1}^{J}(X^{B}_{j\ \text{mean}} - X^{W}_{j\ \text{mean}})\beta^{B}_{j}\right] + [(\beta^{B}_{0} - \beta^{W}_{0}) + \sum_{j=1}^{J}\left(\beta^{B}_{j} - \beta^{W}_{j}\right)X^{W}_{j\ \text{mean}}]$$

The explained difference component

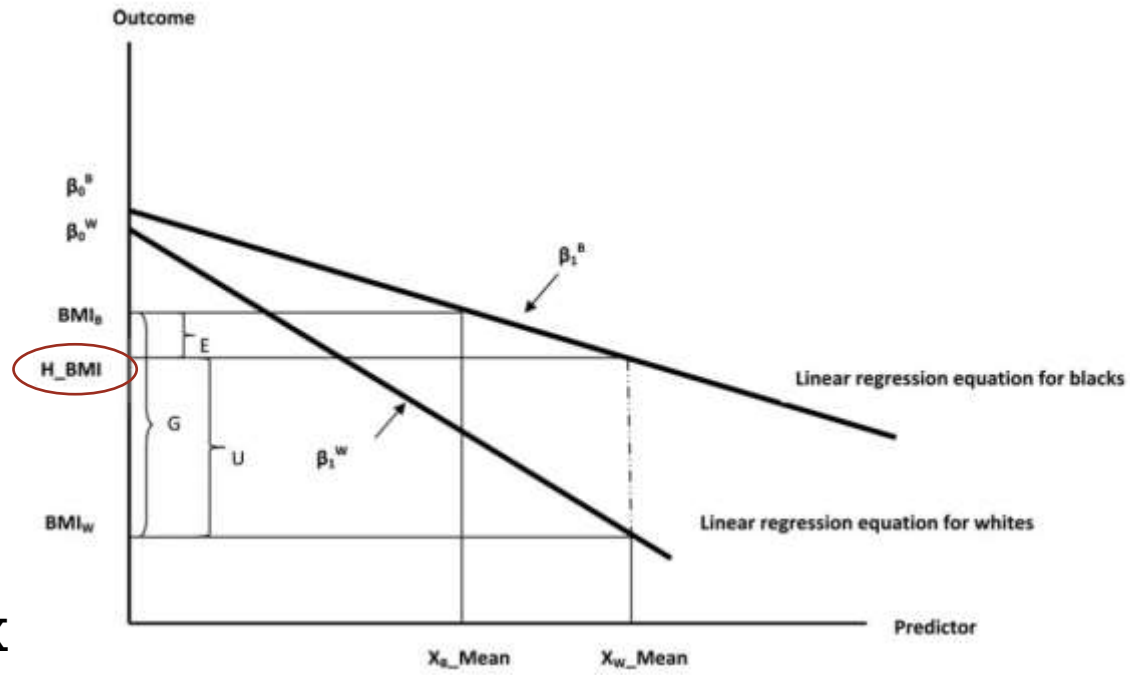
- Portion of the aggregate group differences in BMI attributable to differences in the mean values of the Xj variable(s)
- It represents the amount by which the racial difference in BMI would change in the hypothetical world where, other things equal, black individuals had the same mean levels of Xj as the white individuals

The unexplained component:

- Racial disparity in BMI that would remain even if black individuals had the same mean levels of Xj as the white subjects.
- It also includes the potential interactions

HOW DOES IT WORK? (3/3)

- Assuming one X variable with a negative slope (with Y) and that mean level of X is higher among NH whites than NH blacks
- Both intercepts and slopes differ by racial group
- H_{BMI} is the hypothetical BMI for blacks if they were to have the same level of X as the whites, but their own intercept and slope
  - G denotes the overall racial gap in mean BMI
  - E denotes the portion "explained" by differences in X mean
  - U denotes the "unexplained" portion
- Analogy with Multilevel modelling



THE UNEXPLAINED PORTION

- In the traditional gender-gap in wages literature, this portion was often attributed to employer discrimination against female workers
- But also unmeasured individual or contextual variables
- Systematic measurement error
- Also includes the potential interaction with R/E and included variables
  - Could be specifically decomposed

SEN'S EXAMPLE

TABLE 1 Sample means, regression estimates, and decomposition results for white and black females

Variable[a]	Mean (std dev) White; $N = 7695$	Mean (std dev) Black; $N = 3388$	Coeff White	Coeff Black	Contribution to "explained" gap	95% CI lower	95% CI upper
Age	50.105 (0.346)	44.397 (0.448)	0.007	0.026*	−0.039	(−0.141	0.062)
Education: some college[b]	0.289	0.300	−0.311	−0.666	−0.003	(−0.015	0.008)
Education: college graduate or higher[b]	0.291	0.196	−1.188***	−0.174	0.112	(0.050	0.174)
HH Income $25-$50K per year[c]	0.225	0.192	−0.275	−1.450***	0.009	(−0.018	0.036)
HH Income $50K< per year[c]	0.364	0.163	−1.708***	−2.127***	0.344	(0.177	0.511)
HH Income not reported[c]	0.186	0.142	−2.217***	−1.807***	0.096	(0.035	0.157)
Days of poor mental health last month	5.820 (0.267)	6.089 (0.301)	0.016*	0.012	0.004	(−0.009	0.018)
Has health insurance	0.885	0.749	−0.961**	0.067	0.131	(0.015	0.246)
Employed	0.451	0.465	0.357***	0.032	0.005	(−0.009	0.019)
Married	0.664	0.340	0.720	0.052	−0.234	(−0.411	−0.057)
Children <18 years in household	0.751 (0.022)	1.128 (0.038)	−0.205	−0.176	−0.077	(−0.180	0.026)
Eats five or more servings of fruits and vegetables	0.201	0.177	−0.015	−0.348	0.000	(−0.013	0.014)
Meets recommended moderate PA levels	0.259	0.269	−0.168	−0.696*	−0.002	(−0.008	0.005)
Meets recommended vigorous PA levels	0.138	0.132	−0.019	1.239**	0.000	(−0.004	0.005)
Job is physically demanding	0.120	0.152	−0.110	0.493	−0.004	(−0.029	0.022)
Intercept			28.536***	31.126			
Predicted body mass index	27.028 (6.217)	31.063 (7.740)					
Total explained gap					0.343	(0.082	0.604)
Total unexplained gap					3.732	(3.195	4.268)
Total predicted gap					4.075	(3.628	4.522)

BRFSS data for non-Hispanic Whites and Blacks from Mississippi and Alabama are used in this example. All results are weighted.
[a]Standard deviations are presented for all continuously measured variables.
[b]Reference category is "high school or less."
[c]Reference category is "HH Income <$25K per year."

SOME EXAMPLES

Social Science & Medicine 69 (2009) 793–801

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Inequalities in health care utilization in Spain due to double insurance coverage: An Oaxaca-Ransom decomposition

Mª Luz González Álvarez*, Antonio Clavero Barranquero

University of Málaga, Málaga, Spain

Using Decomposition Analysis to Identify Modifiable Racial Disparities in the Distribution of Blood Pressure in the United States

Sanjay Basu*, Anthony Hong, and Arjumand Siddiqi

IJHPM International Journal of Health Policy and Management

Original Article

Changes in Socio-Economic Inequality in Neonatal Mortality in Iran Between 1995-2000 and 2005-2010: An Oaxaca Decomposition Analysis

Mostafa Amini Rarani[1,2], Arash Rashidian[1*], Ardeshir Khosravi[3], Mohammad Arab[1], Ezatollah Abbasian[4], Esmaeil Khedmati Morasae[5,6]

Medical Care Research and Review
Volume 66 Number 5
October 2009 561-577
© 2009 SAGE Publications
10.1177/1077558709338487
http://mcr.sagepub.com
hosted at
http://online.sagepub.com

Understanding Observed and Unobserved Health Care Access and Utilization Disparities Among U.S. Latino Adults

Arturo Vargas Bustamante
University of California, Los Angeles
Hai Fang
University of Colorado, Denver
John A. Rizzo
Stony Brook University, New York
Alexander N. Ortega
University of California, Los Angeles

ELSEVIER

Social Science & Medicine 62 (2006) 2897–2914

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Health disparities between racial groups in South Africa: A decomposition analysis

Cécile Charasse-Pouélé[a], Martin Fournier[b,*]

[a]*Maître de Conférences, Centre d'Etudes et de Recherches sur le Développement International, 65 bd François Mitterrand, 63 000 Clermont-Ferrand, France*
[b]*CEFC, Room 304, Yu Yuet Lai Building, 43-55 Wyndham Street, Central, Hong Kong*

Available online 23 January 2006

Research

A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP490.

Decomposition Analysis of Black–White Disparities in Birth Outcomes: The Relative Contribution of Air Pollution and Social Factors in California

Tarik Benmarhnia,[1] Jonathan Huang,[2] Rupa Basu,[3] Jun Wu,[4] and Tim A. Bruckner[4]

[1]Department of Family Medicine and Public Health & Scripps Institution of Oceanography, University of California
[2]Institute for Health and Social Policy, McGill University Montreal, Quebec, Canada
[3]Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, California, USA
[4]Program in Public Health, University of California, Irvine, Irvine, California, USA

UNEQUAL EXPOSURE OR UNEQUAL VULNERABILITY?

Original Contribution

Unequal Exposure or Unequal Vulnerability? Contributions of Neighborhood Conditions and Cardiovascular Risk Factors to Socioeconomic Inequality in Incident Cardiovascular Disease in the Multi-Ethnic Study of Atherosclerosis

Mustafa Hussein*, Ana V. Diez Roux, Mahasin S. Mujahid, Theresa A. Hastert, Kiarri N. Kershaw, Alain G. Bertoni, and Ana Baylin

- Hussein et al. decomposed the difference (inequality) in CVD incidence rates between low- and high-SEP groups into contributions of:
  - Differences in covariate means (differential exposure)
  - Differences in CVD risk associated with covariates (differential vulnerability)
- Covariates included:
  - baseline demographics, neighborhood characteristics, and psychosocial, behavioral, and biomedical risk factors.
- "Sequential approach" to deal with confounding

$$\hat{\Delta}_{\text{CVD}} \approx \underbrace{\left(\bar{X}_L - \bar{X}_H\right) \times \hat{\beta}_L}_{\text{Differential Exposure}} + \underbrace{\left(\hat{\beta}_L - \hat{\beta}_H\right) \times \bar{X}_H}_{\text{Differential Vulnerability}} = \hat{\Delta}_E + \hat{\Delta}_V.$$

Table 5. Adjusted Contributions[a] of Psychosocial, Behavioral, and Biomedical Risk Factors to the Observed Socioeconomic Inequality (Low-High) in Incident Cardiovascular Disease (n = 3,729), Multi-Ethnic Study of Atherosclerosis, 2000–2012

Covariate	Differential Exposure[b] (Δ_E) Contributions			Differential Vulnerability[c] (Δ_V) Contributions		
	Absolute, Cases/1,000	95% CI	Relative, %	Absolute, Cases/1,000	95% CI	Relative, %
Aggregate contribution—all factors[d]	2.3[e]	0.4, 4.3	50.7	0.2	−2.6, 3.0	4.0

- Links with the 4-way decomposition

Table 5. Adjusted Contributions[a] of Psychosocial, Behavioral, and Biomedical Risk Factors to the Observed Socioeconomic Inequality (Low-High) in Incident Cardiovascular Disease (n = 3,729), Multi-Ethnic Study of Atherosclerosis, 2000–2012

Covariate	Differential Exposure[b] (Δ_E) Contributions			Differential Vulnerability[c] (Δ_V) Contributions		
	Absolute, Cases/1,000	95% CI	Relative, %	Absolute, Cases/1,000	95% CI	Relative, %
Aggregate contribution—all factors[d]	2.3[e]	0.4, 4.3	50.7	0.2	−2.6, 3.0	4.0
Psychosocial factors[f]						
Aggregate contribution	0.2	−0.5, 0.8	3.4	0.0	−0.5, 0.6	0.6
Lifetime discrimination[g]	−0.1	−0.3, 0.2	−1.4	0.0	−0.2, 0.2	0.6
Chronic stress[g]	0.0	−0.1, 0.2	0.9	0.0	−0.1, 0.1	−0.5
Depressive symptoms[g]	0.2	−0.4, 0.7	4.0	−0.1	−0.6, 0.4	−1.5
Social support[g]	0.0	−0.3, 0.3	−0.1	0.1	−0.3, 0.4	1.9
Behavioral factors[f]						
Aggregate contribution	0.4	−0.6, 1.4	9.5	0.5	−2.0, 2.9	10.1
Uninsured (no health insurance)	0.0	−0.5, 0.5	0.0	−0.1	−0.2, 0.0	−1.6
Pack-years of smoking[g]	0.0	0.0, 0.1	1.0	0.0	−0.1, 0.1	−0.4
Current alcohol consumption	0.5	−0.4, 1.4	11.1	0.4	−2.1, 2.8	8.5
Moderate/vigorous PA[g]	0.0	−0.1, 0.0	−1.0	0.0	−0.1, 0.2	0.4
Body mass index[g,h]	−0.1	−0.3, 0.2	−1.5	0.1	−0.1, 0.4	3.1
Biomedical factors[f]						
Aggregate contribution	1.7[i]	0.6, 2.9	37.7	−0.3	−1.4, 0.7	−6.7
Diabetes[j]	0.5[e]	0.1, 0.8	9.8	0.0	−0.3, 0.2	−0.1
Hypertension[k]	0.9[i]	0.3, 1.5	19.3	−0.3	−1.3, 0.7	−7.0
Total cholesterol[g]	0.1	0.0, 0.2	1.7	0.0	−0.1, 0.1	0.3
Interleukin-6[g]	0.3[l]	0.0, 0.6	5.9	0.1	−0.1, 0.2	1.2
C-reactive protein[g]	0.1	−0.1, 0.2	1.2	0.0	−0.2, 0.1	−1.0

Abbreviations: CI, confidence interval; PA, physical activity; SEP, socioeconomic position.

[a] "Absolute" columns list absolute contributions (number of cases per 1,000 person-years) to inequality by each risk factor through differential exposure and differential vulnerability. "Relative" columns list those contributions as a percentage of the inequality. Estimates were generated in decompositions of the observed, 4.6-extra-case low-high SEP inequality.

[b] Difference in the prevalence of the risk factor between low- and high-SEP groups.

[c] Difference in the association of the risk factor with cardiovascular disease between low- and high-SEP groups.

[d] The overall aggregate contribution ($\Delta_E + \Delta_V$) of all risk factors in model 3 was 2.5 (95% CI: −0.6, 5.7) cases per 1,000 person-years (54.7% of the inequality).

[e] $P < 0.05$.

[f] Contributions of each risk factor in model 3 were adjusted for demographic factors and neighborhood variables, as well as for all other psychosocial, behavioral, and biomedical factors listed.

[g] Variable was specified in decomposition models as a z score (standard deviation units).

[h] Weight (kg)/height (m)^2.

[i] $P < 0.01$.

[j] Fasting glucose concentration ≥126 mg/dL and/or use of insulin/oral hypoglycemic medication (74).

[k] Systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or use of antihypertensive medication (73).

[l] $P < 0.1$.

SENSITIVITY ANALYSES FOR MEDIATION ANALYSES

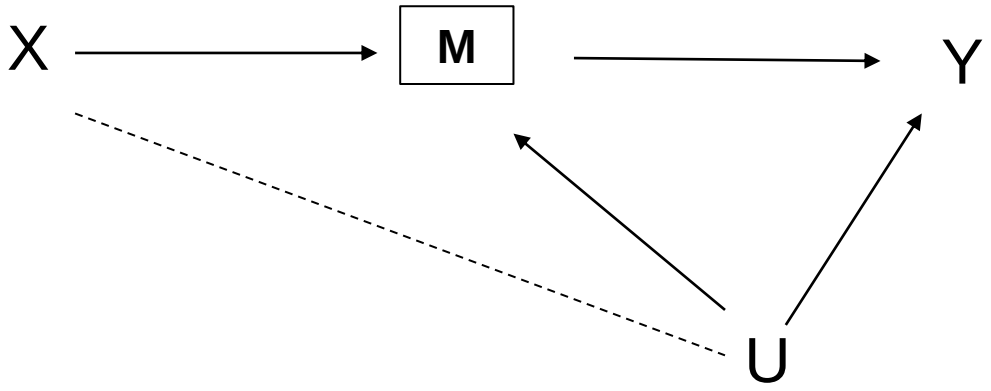
SENSITIVITY ANALYSIS FOR ASSUMPTION VIOLATIONS

- Consider the following assumptions:

1. No residual [unmeasured] exposure-outcome confounding
2. No residual [unmeasured] mediator-outcome confounding
3. No residual [unmeasured] exposure-mediator confounding

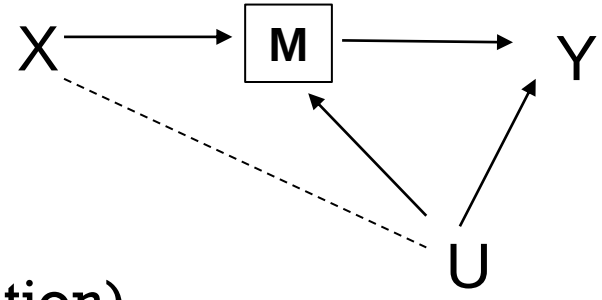
- (1) and (3) are necessary to identify the total effect while (2) is specific to mediation analysis.
- (1) and (3) can be taken into account by randomizing the exposure but randomization will not help (2)

MEDIATOR OUTCOME CONFOUNDING



- C is known and can be conditioned on, but U is unknown or unmeasured
- Conditioning on M (for the CDE) induces an X-U association

BIAS ANALYSIS



- The magnitude of the bias depends on
  - the direct effect of U on Y
  - the magnitude of the collider bias (the induced X-U association)
- 2 sensitivity parameters for the Bounding Factor (BF)
  - Assuming that X and Y are binary (0,1)
- If U is binary:
  - γ is the maximum C-conditioned RR of U onY among the exposed over the M levels
  - λ is the maximum C-conditioned RR of X on U over the M levels

$$\gamma = \max_m \frac{\max_u P(Y=1 \mid X=1,m,c,u)}{\min_u P(Y=1 \mid X=1,m,c,u)} \qquad \lambda = \max_{u,m} \frac{P(u \mid X=1,m,c)}{P(u \mid X=0,m,c)}$$

$$BF = \frac{\gamma\lambda}{\gamma+\lambda-1}$$

Ding et al. Bioetrika 2016;103:483-90

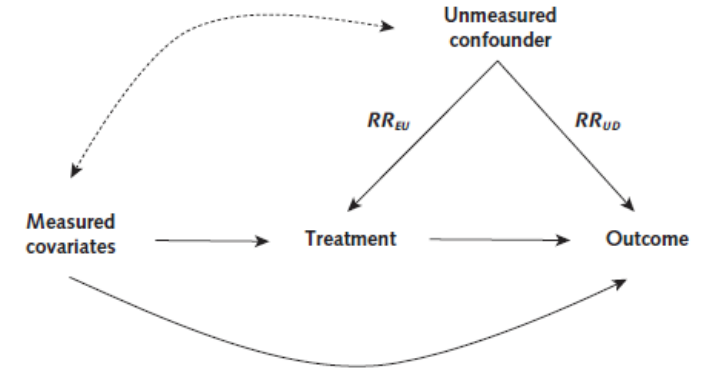
Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Sensitivity Analysis in Observational Research: Introducing the E-Value

Tyler J. VanderWeele, PhD, and Peng Ding, PhD

INTRODUCING THE E-VALUE

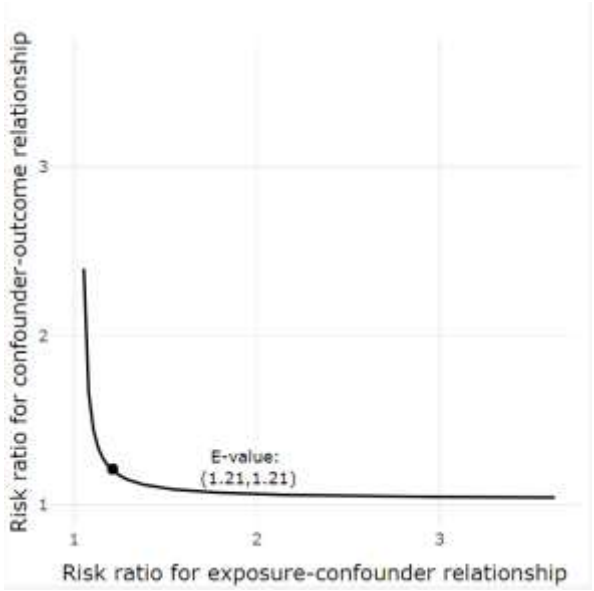
- The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to explain away a treatment–outcome association.
- A large E-value implies that considerable unmeasured confounding would be needed to explain away an effect estimate
- A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate



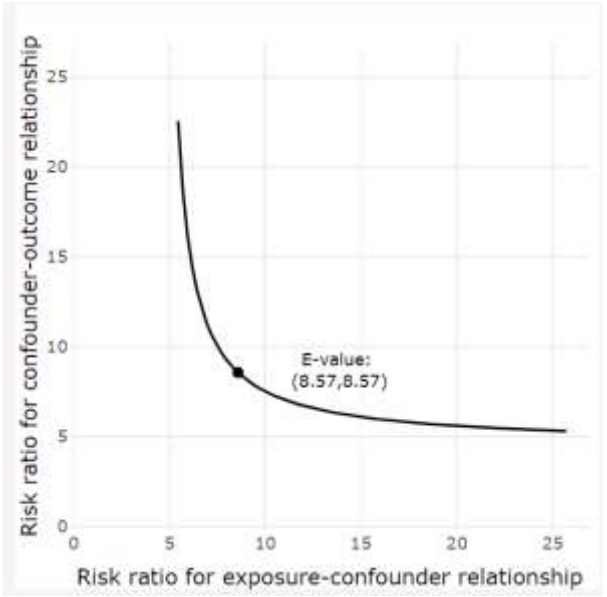
$$\text{E-value} = RR + \text{sqrt}\{RR \times (RR - 1)\}.$$

EXAMPLES

Small effect Size: RR = 1.03 (95%CI: 1.01; 1.05)



Large effect Size: RR = 4.55 (95%CI: 4.21; 4.78)



If one of the two parameters is smaller than the E-value, the other must be larger, as defined by the curve below. All points along the curve define joint relationships that explain away the estimated effect, including points to the right of the curve.